

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/101750/>

**Copyright and reuse:**

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

**Grade decisions: how observers make judgements in  
the Observation of Teaching and Learning**

by

**Graeme Sutherland**

**A thesis submitted in partial fulfilment of the  
requirements for the degree of Doctor of Education**

**University of Warwick, Centre for Education Studies**

**November 2017**

## Contents

<b>List of tables</b>	<b>6</b>
<b>Glossary</b>	<b>8</b>
<b>Acknowledgements and Declaration</b>	<b>9</b>
<b>Abstract</b>	<b>10</b>
<b>Chapter 1 Introduction</b>	<b>11</b>
1.1 The aim of the research	11
1.2 Background/setting	12
1.3 Personal history	13
1.4 Researcher perspective	14
1.5 Why does this research matter?	16
1.6 The focus of this thesis	18
1.7 Overview of thesis	20
<b>Chapter 2 Literature Review</b>	<b>22</b>
2.1 Introduction	22
2.2 Review method	22
2.3 History, development and current practice of OTL	24
2.3.1 Further Education and Adult and Community Education policy and context	25
2.3.2 History and development of OTL	26
2.3.3 Current practice and purposes of OTL	27
2.4 Participants' experience of OTL	29
2.4.1 Observees' experience of OTL	30
2.4.2 Observers' beliefs about OTL	32
2.4.3 The relationship between observer and observee	33
2.5 OTL decision-making	34
2.5.1 Criteria	36
2.5.2 Expert judgement	40
2.6 Wisdom of Practice	43
2.7 'Effective Teaching'	49
2.8 Outcome measures of 'effective teaching'	56
2.9 Summary	61
<b>Chapter 3 Research Design</b>	<b>63</b>
3.1 Introduction	63
3.2 Ontological and epistemological positions	64
3.3 Single-item questionnaires	68
3.4 Interviews	71
3.5 Analysis of secondary data	76
3.5.1 Observation reports	77
3.5.2 Accreditation and achievement data	80
3.5.3 Attendance data	81
3.5.4 Judgement-criteria	81
3.6 Ethical considerations	83
3.7 Summary of methods	86
	89

<b>Chapter 4</b>	<b>Ethnographic Reflections</b>	<b>94</b>
4.1	Introduction	94
4.2	Auto-ethnography	95
4.2.1	Experiences of being observed	96
4.2.2	Experiences of carrying out observations	105
4.3	Observing Participation	109
4.3.1	Co-observing with an experienced observer	111
4.3.2	Co-observing with an inexperienced observer	113
4.4	Summary	118
<b>Chapter 5</b>	<b>Findings of the Research</b>	<b>120</b>
5.1	Introduction	120
5.2	RQ1: What do observers believe about the OTL process?	121
5.2.1	All the observers believed that OTL was useful and important	121
5.2.2	All observers identified a QA purpose	122
5.2.3	All observers identified a CPD purpose	122
5.2.4	Most observers identified an Ofsted inspection purpose	123
5.2.5	All observers were critical of elements of the grading process	124
5.2.6	All observers had concerns about the effects of institutional targets and managerial priorities on the efficacy of OTL	125
5.2.7	Reflections on findings of RQ1	126
5.3	RQ2: How do observers explain their OTL decision-making?	127
5.3.1	All the observers cited intuitive/holistic decision-making	127
5.3.2	All observers also cited criteria-based decision-making	129
5.3.3	All observers believed their OTL decisions were affected by their relationships with observees	131
5.3.4	All observers cited their own professional experience as a basis for decision-making	133
5.3.5	All observers believed their OTL decisions were affected by external factors and considerations	133
5.3.6	All observers expressed doubts about their OTL decisions	134
5.3.7	Reflections on the findings of RQ2	136
5.4	RQ3: What elements of 'effective teaching' do observers focus on in evidencing their OTL decisions?	137
5.4.1	All observers questioned cited judgement-criteria as the main evidence for their OTL decisions	138
5.4.2	Most frequently occurring judgement-criteria in the observation reports	139
5.4.3	Evidence from the data suggested that observers operated a 'hierarchy of criteria' in which some judgement-criteria are valued less highly than others	142
5.4.4	There was a clear relationship between OTL grade and incidence of frequently cited judgement-criteria as strengths and improvements	143

5.4.5	Each OTL grade was associated with a different profile of specific judgement-criteria	144
5.4.6	Reflection on the findings of RQ3	147
5.5	RQ4: How do observers' judgements relate to other potential outcome measures of 'effective teaching'?	148
5.5.1	All observers questioned accreditation an important OTL focus	148
5.5.2	Achievement and accreditation rates had no observable relationship with OTL grades <i>overall</i> (but a difference was observed at the boundary of the acceptable OTL grades)	150
5.5.3	Attendance rates had no observable relationship with OTL grades	151
5.5.4	Separate outcomes were associated with different profiles of specific judgement-criteria	153
5.5.5	Reflection on the findings of RQ4	155
5.6	Summary of key findings	156
<b>Chapter 6</b>	<b>Discussion</b>	<b>158</b>
6.1	Introduction	158
6.2	Discussion of findings by research question	159
6.2.1	What do observers believe about the OTL process?	159
6.2.2	How do observers explain their OTL decision-making?	160
6.2.3	What elements of 'effective teaching' do observers focus on in evidencing their OTL decisions?	162
6.2.4	How do observers' judgements relate to other potential outcome measures of 'effective teaching'?	165
6.3	Profiles of judgement-criteria	167
6.4	Consistency between findings	168
6.5	The relationship between OTL decision-making and the concept of Wisdom of Practice	172
6.5.1	Introduction	172
6.5.2	Evidence for Wisdom of Practice	173
6.6	"How do observers make judgements in the observation of Teaching and Learning?" - A model	177
6.6.1	The model	180
6.6.2	Critique of the model	186
<b>Chapter 7</b>	<b>Conclusion</b>	<b>187</b>
7.1	Introduction	187
7.2	Overview of the research	187
7.3	Summary of key findings	188
7.4	Strengths	190
7.4.1	Focus on the observer's OTL experience	190
7.4.2	Ethnography	192
7.4.3	Analysis of secondary data	193
7.4.4	The role of Wisdom of Practice in the OTL process	195
7.4.5	Model of the OTL judgement process	196
7.5	Limitations	196
7.6	Implications and recommendations	198
7.6.1	Theoretical	198

7.6.2	Practical	199
7.6.3	The future for OTL?	200
<b>References</b>		<b>203</b>
<b>Appendices</b>		<b>216</b>
Appendix 1	Judgement-criteria cited in observation reports	217
Appendix 2	Auto-ethnography questionnaire items	218
Appendix 3	Interview schedule	220
Appendix 4	Elements of 'effective teaching' from literature sources	221
Appendix 5	Literature sources for elements of 'effective teaching'	222
Appendix 6	Ethical approval	224
Appendix 7	Observee questionnaire results for comparison purpose	226

## List of Tables and Figures

Table 1.1	Summary of observation grades	13
Table 2.1	‘Teaching strategies commonly cited as influencing judgments’ (from Strong et al., 2011, p374)	40
Table 2.2	‘Cognitive operations that influence judgements of human behaviour’ (adapted from Strong et al. (2011, p369))	42
Table 2.3	Comparison of Shulman’s ‘Categories of the knowledge base’ (1987) and Lunenberg and Korthagen’s ‘Experience, theory, and practical wisdom in teaching’ (2009)	44
Fig. 2.4	The Triangular relationship between practical wisdom, theory, and experience. (Lunenberg and Korthagen, 2009, p229, fig.1)	45
Table 2.5	Most frequently occurring elements of ‘effective teaching’ in the literature	54
Table 2.6	Explanation of some key elements of ‘effective teaching’	56
Table 3.1	Example of a typical entry for an individual observation report	78
Table 3.2	Example of a typical accreditation and achievement data entry for a small class	80
Table 3.3	Examples of conversion from verbatim judgement to final judgement-criterion	85
Table 3.4	Method details	90
Table 3.5	Methods used by Research Question	91
Table 4.1	Details of observations 1 and 2	96
Table 4.2	Reflections on observer’s judgement-criteria (Observation 1)	103
Table 4.3	Reflections on observer’s judgement-criteria (Observation 2)	104
Table 4.4	Details of observation 3	106
Table 4.5	Details of observations 4 and 5	110
Table 5.1	Top 20 OTL judgement-criteria cited in observation reports overall by frequency	139
Table 5.2	Judgement-criteria and some trends in the data	140
Table 5.3	Frequency of judgements by grade (as strengths and improvements)	144
Table 5.4	Most frequent 5 strengths and improvements by grades	145
Table 5.5	Average OTL grades against average accreditation and achievement rates	150
Table 5.6	Average OTL grades compared with average attendance rates	152
Table 5.7	Profiles of judgement-criteria as strengths for high accreditation rate, attendance rate and grade 1 observations (for comparison)	153
Table 6.1	Characteristics of the Wisdom of Practice compared to findings	174
Table 6.2	Comparing the ranking of most frequent 20 judgement-criteria in observation reports with frequency of occurrence in sample of sources from literature review.	176

Fig. 6.3	The relationship between experience, theory, knowledge of overall context and ‘Wisdom of Practice’ (adapted from Lunenberg and Korthagen, 2009, p229, fig.1)	<b>179</b>
Fig. 6.4	“How do observers make judgements in the Observation of Teaching and Learning?” A model. (Partly suggested by Lunenberg and Korthagen (2009, p229, fig.1))	<b>185</b>
Table A1.1	Full list of judgement-criteria cited in observation reports	<b>217</b>
Table A4.1	Full list of elements of ‘effective teaching’ from literature sources	<b>221</b>
Table A5.1	Sources accessed in the survey of literature on the subject of ‘effective teaching’	<b>223</b>
Table A7.1	Comparison between the 20 most frequent elements of ‘effective teaching’ from the observee questionnaire with most frequent judgement-criteria in the observation reports	<b>226</b>



## Glossary

Acc	Accreditation
Ach	Achievement
ACE	Adult and Community Education sector
AES	Adult Education Service
AfL	Assessment for Learning
BME	British Black and Minority Ethnic
CLASS	Classroom Assessment Scoring System
CPD	Continuing Professional Development
DfES	Department for Education and Skills
E&D	Equality and Diversity
ES	Essential Skills
ESOL	English for Speakers of Other Languages
ETF	Education and Training Foundation
FE	Further Education
frq	Frequency
GCSE	General Certificate of Secondary Education
HE	Higher Education
ILP	Individual Learning Plan
Improvements	‘Areas for Improvement’ judgement-criteria
imps	Improvements
ICT	Information and Communications Technology
LEA	Local Education Authority
LSC	Learning and Skills Council
Newbold	Pseudonym for the Adult Education Service setting of this study
observer	Manager with observer-role
occ	Occurrence
Ofsted	Office for Standards in Education
OTL	Observation of Teaching and Learning
PRESET	Pre-Service Education and Training
QA	Quality Assurance
QCF	The Qualifications and Credit Framework
QIP	Quality Improvement Plan
RQ	Research Question
SMT	Senior Management Team
Strengths	‘Strength’ judgement-criteria
TLC	Transforming Learning Cultures in Further Education
TTA	Teacher Training Agency
UCU	The University and College Union
VLE	Virtual Learning Environment

## **Acknowledgements**

I would like to take this opportunity to express gratitude to several people who were essential to the completion of this thesis.

Firstly, I would like to thank my two supervisors at Warwick, Dr Michael Hammond and Dr Justine Mercer, for their guidance and constructive criticism throughout. They have kept me on my toes with just the correct balance of positive feedback and judicious challenge I required. It has been a pleasure to work with them.

I am also grateful to the management and staff of the Adult Education Service referred to below as 'Newbold', the setting for this research, for their openness and cooperation in allowing me to collect my data. I would especially like to thank those colleagues who gave up their time to take part in interviews and questionnaires, and allow me to come into their classrooms.

Finally, I want to acknowledge the invaluable support of my family in what has been a long hard trail for all of us. My love and thanks to my children, William and Daisy, for being so understanding about all the time I've had to miss being with them and yet always remaining so positive and encouraging. And to my beloved wife, Kathryn, most important of all - without her *nothing* would have been possible.

## **Declaration**

I declare that this thesis is solely my own work and has not been submitted for a degree at another university.

*Graeme Sutherland*

## **Abstract**

This research looked at how observers make judgements in the Observation of Teaching and Learning (OTL). Until recently, graded OTL has been the dominant model for assessing teaching quality across all sectors of education in the UK. From Ofsted inspection to routine in-service appraisal, teachers have become familiar, if not comfortable, with being watched and judged. Although the outcomes of these observations have high-stakes consequences for practitioners, and their institutions, the judgement process of the observer has been an under-developed area of research, and questions of which elements of theory, experience and practical knowledge they access in doing so, are areas yet to be fully explored.

This research constituted a case study focussed on the process of OTL in a city-wide Adult Education service provider in England, attempting to understand and clarify the opaque process utilising methods including: interviews and questionnaires with 14 experienced classroom observers, analysis of data from 137 observation reports, and accreditation results and attendance rates pertaining to over 230 classes. Data were complemented by a rich background of ethnographic and auto-ethnographic reflection to provide context and to make the experience of OTL relatable for the reader.

The findings suggested that the concept of Wisdom of Practice was at the heart of the OTL process and that observers appeared to reach their initial judgements holistically based on a largely intuitive and automatic access to a wealth of theory, experience and practical knowledge. Observers then appeared to make use of a range of elements to refine and justify their judgements: chiefly judgement-criteria arising from a range of sources, but also contextual factors including the observer's relationship with the observee.

The analysis and discussion of these findings has led to the proposal of a theoretical model of the OTL judgement process, which clearly demonstrates the complex strands of OTL decision-making. Ultimately, the findings support a claim for the continued usefulness of OTL with an increased emphasis on Continued Professional Development. There is an appreciation of the authenticity of the measure, once grading has been removed, based upon the Wisdom of Practice of the experienced observer.

## Chapter 1 – Introduction

### 1.1 The aim of the research

My overarching aim, in this study, was to address the question of how classroom observers arrived at their judgements during the process of the Observation of Teaching and Learning (OTL). I wanted to shed light on the decision-making process at the heart of this practice, not only because OTL carries high-stakes consequences for practitioners and their institutions alike, but because of what a closer examination of the observers' judgements might reveal about how they define and conceptualise 'effective teaching'.

Throughout the UK education system, the task of making official judgements about teacher effectiveness falls on two distinct sets of shoulders: Ofsted, tasked by the Government to regulate educational institutions; and practitioner-managers (observers) within those institutions, carrying out classroom observation within both their Quality Assurance (QA) and Continuing Professional Development (CPD) procedures - distinct purposes which may or may not be compatible. These two types of classroom observer, the external inspector and the in-house manager-observer, were clearly very different in terms of their purposes, job role and relationships with those they observe. The focus of this study was on the latter, the in-house observers, and how they reached their OTL judgements.

In order to judge whether teaching is effective, observers must already have a conception, however tacit, of what 'effective teaching' is. I was interested in the bases of this conception, especially the extent to which it could be located within the framework of the "personal knowledge of practitioners" (Hawkins, 1966, p3), a shared experience that Shulman (1983, p11), *after* Hawkins, refers to as the 'Wisdom of

Practice’. My aim was to explore the observers’ experience of OTL and to find out what observers believed, felt and understood about observation; to find out how their judgements were evidenced and explained; and how they related to other measures of ‘effective teaching’.

## **1.2 Background/setting**

The setting for this study was a local government Adult Education Service (AES) in a medium-sized city in England. For the purpose of participant anonymity, I have chosen to refer to this institution as ‘Newbold AES’. Total enrolments during 2014-16 were over 25,000, with courses in 11 areas of learning, comprising accredited and non-accredited programmes and including vocational qualifications, apprenticeships, traineeships, Family Learning, and Essential Skills (ES). The service used 24 venues for its provision including schools, community centres, libraries and voluntary organisations. About 75% of learners were female; over 40% self-identified as black and minority ethnic (BME); 18% declared a disability or learning difficulty.

Newbold’s Essential Skills (ES) department was typical of that within the Adult and Community Education sector (ACE), offering ES courses standardised to the Adult Core Curriculum (DfES, 2001), and providing accredited English, Maths and English as a Second or Other Language (ESOL) courses for adults (19+). Qualifications offered included Functional Skills at Levels 1 & 2 and QCF awards at entry level.

Newbold’s mission statement (2010) expressed a learner-centred, community-based focus, with an ethos concerned with the “celebration and promotion of active learning”; “personal growth”; and “equality of opportunity”. However, at the time of writing, ES provision was funded primarily on the basis of accreditation outcomes, part of a shift,

within the last ten years, to achievement-focused targets and assessment with "accountability purposes" for which the "new emphasis ... is that of meeting targets which have been imposed by those supplying the funding" (Stobart, 2008, p118).

During the period covered, all Newbold practitioners were annually observed and awarded a grade for their teaching following the then current Ofsted grading system (although not based on any given observation schedule):

Grade 1	Outstanding
Grade 2	Good
Grade 3	Requires Improvement
Grade 4	Inadequate

*Table 1.1: Summary of observation grades*

### **1.3 Personal history**

My first experiences of OTL were during postgraduate teacher training (1991-2) during which I was routinely observed by teacher-trainers in an atmosphere of caring, critically-constructive support. This became, for me, the model of highly effective OTL – a genuinely educative experience in which improvements were identified and facilitated within an unthreatening and non-judgemental framework.

In 1992, I began teaching in a large comprehensive school in the north of England. The prospect of the school's first Ofsted inspection initiated the introduction of a formal OTL policy. This was presented as focussing on teachers' CPD, but was conducted in

an atmosphere of tension wrought by the looming Ofsted visit. This was my first experience of witnessing practitioner resentment of an OTL regime.

I have worked within adult education since 1996, from 2003 at Newbold AES, teaching basic ES and GCSE. In 2008, I became a Programme Manager, which involved responsibility for carrying out classroom observations.

Although, throughout my time at Newbold, I have never achieved less than Grade 2 in annual observations, with the majority being Grade 1, it was still the case that like many practitioners, I found the process as described by Lasagabaster and Sierra (2011): “stressful and intimidating” provoking “uneasiness, nervousness, and tension” (p450).

#### **1.4 Researcher perspective**

Arising from, and being situated in, my own professional practice, my perspective might best be considered as that of an *insider*. Insider research has been defined as that being:

“...undertaken by people who, before they begin to research, already have an attachment to, or involvement with, the institutions or social groups in, or on, which their investigations are based.”

(Sikes and Potts, 2008, p3)

Insider research has many practical advantages. It also has a potential claim to a privileged order of validity, or what might be better termed, *authenticity*. Practical advantages included easier access to subjects for interview, survey and observation and to data within Newbold’s files. I was also able to address my own personal experience of OTL as a source of illustration and in order to make the experience of OTL relatable

to the reader. In interviews, I found that I had a privileged status with my subjects, “considerable credibility and rapport” (Mercer, 2007, p7), based on 13 years of familiarity, making it easier to put them at ease, to arrive at mutual trust, to facilitate dialogue – to “engender a greater level of candour than would otherwise be the case” (ibid).

There were, however, a number of concerns arising from my status as an insider. For example, there was the potential for collusion in these interviews, from a sense of shared ‘understanding’ and ‘experience’, which may have resulted in constructing a context in which certain emergent meanings affected what was said or not said – and how I interpreted the material. There was also the question of the effect of my relationships with individual subjects: “the power relationships within which the researcher and the researched co-exist [and] the personalities of the researcher and specific informants” are important factors (Mercer, 2007, p4). I was “closer” to some of my interviewees than to others; I had managed some, and have been managed by others. The extent to which these individual differences affected the data would be impossible to determine – although data arising from interviews and other primary sources in this research, should be considered inherently subjective in the first place. Inevitably, the responses discussed here were as human, constructed and mutable as the subject matter they were focussed upon.

In terms of the claim of *authenticity*, ultimately, the key strength of insider research is the researcher’s privileged appreciation of context. I was able to utilize my experience of observing and being observed in order to understand the data, what Sikes (2008) calls “a privileged understanding of the process” (p154). The experience of insider-status becomes a lens through which to bring the research into focus. A potential drawback was the inevitable subjectivity and potential lack of detachment – therefore, following



Smyth and Holian (2008, p36), my focus has been primarily “concerned with relevance, usefulness, [and] resonance” rather than that which needs to be “proven, validated or replicated”, as a more positivistic approach might demand. This should be understood and taken into consideration by those reading my work. However, I hoped that this thesis would provide ample evidence that I “regard as essential: rigour, robustness, transparency of process and method, systematic and internally consistent approaches to data gathering and analysis, a clear chain of evidence and ethical practices” (ibid).

Although I have endeavoured throughout to maintain a distance between my views and feelings and the collation and interpretation of results, this separation was essentially notional in nature. My experiences, and their effects on my outlook, necessarily coloured my approach and interpretation. It therefore seemed appropriate and useful to incorporate these into this work, forming the basis for the ethnographic reflections detailed in **Chapter 4**.

## **1.5 Why does this research matter?**

Although, during the period of data-collection, graded OTL was common practice throughout most sectors of education in England and Wales, at the time of writing, this practice was under review, with many institutions considering abandoning the application of grades.

Coe et al (2014,) stated that using the “best classroom observation ratings” available, “we would get it right about 60% of the time, compared with the 50% we would get by just tossing a coin”, concluding that OTL judgements “need to be used with considerable caution” and suggesting an approach based on “multiple measures” (p3). Not only then is graded OTL, potentially, an ‘inefficient’ measure, *in these terms*, it was

also one that inflicted considerable “anxiety, stress and insecurity [on] members of the profession” (O’Leary, 2006, p196).

In the light of such research, Ofsted (2015) have expressed serious doubts concerning the efficacy of OTL, stating that they will no longer “grade the quality of teaching”, and classroom observation would be “supplemented by a range of other evidence” (p45). At the time of writing, Newbold was actively considering following Ofsted in abolishing grading.

It was, therefore, timely for research to be carried out in order to record, before its possible demise, what the practice of grading had to tell us about the way in which the concept of ‘effective teaching’ had been judged by observers. If grading was to be replaced by a more formative process, that might make decision-making even more difficult to penetrate. Whatever the negative effects of grading, the process did at least involve a ‘clear’ externally expressed judgement which might be more open to analysis.

It was also important to recognise that OTL had been the only assessment of ‘effective teaching’ based on the direct judgement and feedback of experienced practitioners. Alternatives to OTL necessarily involve indirect measures: particularly, assessment and accreditation outcomes. Such ‘Performance Indicators’ have the advantage, perhaps, of being superficially more objective than the judgement of a single observer. However, it could be argued that teaching is only one factor in producing such outcomes, and that not all the desired outcomes of teaching can be measured by the results of tests.

Perhaps this was why OTL continued to be utilised by Ofsted, even as grading was abandoned. Coe et al. (2014, p2), despite asserting that “student progress is the yardstick by which teacher quality should be assessed”, nevertheless recommended that the best approach to monitoring ‘effective teaching’ was to “triangulate” the various

methods - including OTL. As Norris (1993) suggested, “the aims and purposes of education are many and its values various” and are “not susceptible to simple forms of measurement” (p36), that there was more to ‘effective teaching’ than the delivery of learning targets and certificates.

Part of my purpose here, therefore, has been to explore the potential ‘usefulness’ of OTL, in terms of how much it was based on and reflected the expertise and knowledge of the observers who carried it out. I believed it would be important to make explicit the way observers brought their Wisdom of Practice to bear on their decision-making, and to provide evidence for the conclusion that OTL represented a unique and valuable resource.

## **1.6 The focus of this thesis**

This thesis takes the form of a case study of OTL within Newbold’s essential skills department. It presents qualitative findings based on interviews and questionnaires reporting what observers said, and believed, about the judgement processes behind their OTL decision-making. Analysis of secondary data was used to explore the relationship between actual grade-decisions and the judgement-criteria used to evidence them.

My overarching aim, in this study, was to address the question:

**‘How do observers make judgements in the Observation of Teaching and Learning?’**

This question led on to the following key research questions (RQs):

**I      What do observers believe about the OTL process?**

The heart of OTL is observer judgement: observers' decisions; how they attempted to evidence these decisions; and why these kinds of judgements might be made.

**II     How do observers explain their OTL decision-making?**

I was interested in how observers rationalised their decision-making, what features of teaching they prioritised, and what, if anything, made decision-making difficult or stressful. A range of methods was used to collect this data: including face-to-face interviews, and questionnaires.

I also wanted to address a **subsidiary question**: Was there evidence that the concept of Wisdom of Practice was involved in OTL decision-making?

**III    What elements of 'effective teaching' do observers focus on in evidencing their OTL decisions?**

I undertook an analysis of the judgement-criteria cited by observers in interviews and questionnaires, and as recorded in observation reports, in order to shed light on the types of evidence used to support grade-decisions: their possible meanings, origins and potential relation to the concept of Wisdom of Practice.

A comparison was undertaken between the judgement-criteria cited by practitioners, and elements of 'effective teaching' found in the literature.

#### **IV      How do observers' judgements relate to other potential outcome measures of 'effective teaching'?**

From a common-sense perspective, rates of learner achievement and accreditation and rates of learner attendance on a particular practitioner's course might reasonably be presumed indicative of the effectiveness of the teaching. However, these relationships have not been proved, and appear far from straightforward. Analysis was undertaken to investigate the relationship between observers' OTL judgements and data relating to other potential outcome measures of 'effective teaching'.

A further purpose was to discover the extent to which OTL represented a unique form of assessment, focussing on different, but equally important, aspects of 'effective teaching'.

#### **1.7      Overview of thesis**

This introductory chapter has outlined the aims; provided background information about the researcher and the research environment; and introduced the research questions and focus.

**Chapter 2** offers a review of the relevant literature, including that referring to: the history, development and current practice and purpose of OTL; participants' experience of OTL; studies on key areas such as decision-making, criteria, expert judgement and the concept of 'Wisdom of Practice'. There is a summary of the literature referring to 'effective teaching', and on alternative outcome measures of 'effective teaching'.

**Chapter 3** outlines the ontological and epistemological positions behind this research and offers details of the research methods used to collect the data, including: the use of questionnaires, the conduct of interviews, and the collection and analysis of secondary data. It also discusses ethical considerations arising from the collection of data.

**Chapter 4** presents a series of ethnographic and auto-ethnographic reflections based on personal experiences of being observed, of carrying out observations. It features subjective analysis of the meanings behind both the preparation for being observed, and the feedback received. The specific methodological issues involved are also discussed.

**Chapter 5** presents the findings of the primary and secondary data collection, organised in reference to the research questions.

**Chapter 6** offers a critical discussion of the key findings, with reference to the literature. The relationship of OTL decision-making to the concept of Wisdom of Practice is discussed and evaluated. A theoretical model of the OTL judgement process is presented.

**Chapter 7** presents conclusions; it summarises the findings from all sources and offer a critical commentary. It includes an evaluation of the strengths and weaknesses of the research design, and a personal statement of how my understanding has been affected. It ends with the discussion of the implications of the findings for both practical implementation and future theoretical work.

Finally, there is a complete list of the literature referenced in this thesis, and a number of appendices.

## **Chapter 2 - Literature Review**

### **2.1 Introduction**

In this chapter, I have provided an overview of the history and current practice of OTL, and discussed relevant concepts with reference to the literature within the field. The focus of this thesis, and the RQs addressed within it, has been informed by the discovery of certain gaps apparent in the current literature, which will be discussed in their respective sections. Therefore, following a brief explanation of my review methods, I have structured this chapter around these RQs, inasmuch as a reading of the literature provided a background and context for the further exploration of these questions; and offered an explanation and discussion of the key concepts involved.

### **2.2 Review method**

I conducted library and on-line searches; the latter using the Google search engine and, especially, Google Scholar due to the ease of linking from individual search results to documents held by the university library. I was able to identify various themes within the literature, including various permutations and elaborations of the following:

- Observation of teaching and learning
- Classroom observation
- Theory of criteria
- Expert decision-making
- Definitions of ‘effective teaching’
- Elements of ‘effective teaching’
- Wisdom of Practice

These provided the focus of my reading, and were used to develop the sections below based around my four RQs.

I needed to be judicial in my approach to the material my searches were unearthing, in order to decide what was relevant. The fact that most of the studies mentioned below are connected with each other through a web of references and citations at least helped to ensure an internal consistency. I excluded a great deal of what appeared to be guidance or policy material, as I was looking specifically for academic peer-reviewed studies.

Accessing literature as a result of these searches, my reading then ‘snowballed’, moving from references within texts to other studies cited within, and occasionally beyond, the field. In this way, I was able to follow trains of thought and perspective, and also source material that I would not have otherwise encountered in my initial searches. It should be noted that the range of material available on the subject of OTL was not broad in itself, especially beyond the sub-topic of the observees’ subjective experience of OTL, and it was unlikely that I would have missed out a significant area of the literature.

I considered material originating in various countries, especially on the subject of ‘effective teaching’. I have focussed mostly on UK work pertaining to the specific context of OTL as being most relevant to my own setting, but took a wider approach when considering the discussion of theory. Much of the work on Wisdom of Practice, for example, has been American in origin and it was important to include it. I considered overseas conceptual studies more relevant despite their origin as these are arguably less reliant on practical context and setting than more empirical work.



O’Leary (2011) concluded that there was “a noticeable gap in the literature of empirical studies exploring the perceptions and experiences of in-service tutors, especially those in FE” (p20), and this was not a situation that appeared to have improved at the time of writing, O’Leary’s own sustained contributions aside. There was very little in the literature on observers’ views of the process, and even less on how observers reach their OTL decisions

### **2.3 History, development and current practice of OTL**

This section relates to RQ 1: What do observers believe about the OTL process? It provides background discussion of the history, development and current practice of OTL, against which observers operate and which informs their beliefs. It concludes by discussing some of the findings from the literature looking at practitioners’ subjective experiences of OTL.

There was little within the literature specifically concerning how observers experience the OTL procedure, and specifically how they reach their OTL decisions. This seemed to be an important omission, as observers’ subjective experience of OTL may well impact on their decision-making.

This section begins with an overview of the policy and context of the post-16 sector that forms the setting for this research.

### **2.3.1 Further Education and Adult and Community Education policy and context**

The specific setting of this research is within the relatively under-researched Adult and Community Education (ACE) sector. Related to, but separate from the much larger Further Education (FE) sector that caters predominantly for 16-19 year olds, ACE providers such as Newbold mainly draw their learners from the post-19 adult population. The division between the two sectors was made more concrete when, in 1992, the Further and Higher Education Act removed FE colleges from local authority control. Most ACE providers, Newbold included, remain, at the time of writing, under local government remit.

Contemporary adult education provision has its roots in the Industrial Revolution, when the “growing concentration of the population in towns” provided the impetus and necessity to extend “the opportunity for ordinary working people to gain instruction” (Hillage et al., 2000, p25). After the First World War, local government ‘night schools’ began to provide “mass adult education opportunities for people to gain qualifications” (ibid).

The creation of the Manpower Services Commission in the 1970s encouraged the development of the sector, recognising that “Britain could not meet its skills needs solely by focussing on the preparation of young people” (ibid, p27). Various initiatives emerged throughout the 1980s, including Employment Training and RESTART .

From 1997, with a new Labour government, there was a significant change of emphasis particularly with reference to ACE. During the late 1990s, the sector saw an “increase in funding for post-school education and the ‘ear-marking’ of considerable sums for adult learners” (Hillage et al., 2000, p31). This initiative was focussed on narrowing the

perceived skills gap in the adult working age population and to “widen participation more generally” (ibid, p30).

Nevertheless, possibly the most important contribution to the literature on post-16 education, the Transforming Learning Cultures in Further Education project (TLC), conducted between 2001 and 2005, found that FE provision was “pressured and destabilised by a combination of inadequate and unstable funding” (James et al., 2007, p145). The report found that these pressures were increasing, and that the consequences were demonstrated within a culture of practitioners working to mitigate the effects of external factors on their delivery, for example: “drastic reductions in class contact time ... and tensions between inclusion and high achievement rates” (ibid, p146). It was recognised that pressure to improve the effectiveness of teaching in the FE sector was “primarily externally driven ... by factors other than the nature of teaching and learning”: wider social and economic concerns alongside the need to be “cost-effective” (ibid).

### **2.3.2 History and development of OTL**

School inspections were first introduced in England and Wales in 1833, although there was comparatively little formal OTL in schools from the 1950s until the end of the 1970s (Mercer et al., 2010, p141). OTL as part of the compulsory appraisal of teachers emerged from the 1986 Education Act, and the 1991 Education (School Teacher Appraisal) regulation, with Ofsted (1996) suggesting that appraisal should “focus more sharply on classroom performance” (p141-p142). It subsequently became the standard experience for school teachers in England and Wales to face classroom observation on an annual basis.

Mandatory OTL was not introduced into the post-compulsory education sector, including Newbold, until 2003 when the Learning and Skills Council (LSC) “directed all providers of post-compulsory education and training to implement systems and policies of observation of teaching and learning” (Burnell, 2016, p231). Individual institutions were allowed to devise their own OTL policies. This represented a radical change of approach:

“Never before, in the history of post-compulsory education, had there been a structured and formal system of OTL ... a formal system of observations with grading, training and development, and possibly sanctions if grades were inadequate”

(Burnell, 2016, p232).

Ofsted served as both a model and a focus for OTL approaches within institutions, seeking to “replicate the Ofsted approach” as part of a “‘mock Ofsted’ in order to prepare staff for the ‘real thing’” (O’Leary, 2013, p708).

### **2.3.3 Current practice and purposes of OTL**

Turner and Clift (1988, p59) identified two fundamental purposes of teacher appraisal:

- a) *Formative appraisal* concerned with professional development, the improvement of practice by identifying strengths, weaknesses, needs and interests.
- b) *Summative appraisal* concerned with the selection, promotion, redeployment and dismissal of teachers

In practice, the picture was complicated by the fact that in many institutions, such as Newbold, the two functions of appraisal were addressed simultaneously by means of

the same method: OTL. There was the risk that the formative function may be seen as being wholly subsumed by the summative. Another factor was that institutions regarded OTL as a strategy for “preparing for impending inspection by Ofsted” (Burnell, 2016, p227).

Wragg (1999), whilst acknowledging that observation has other uses within Initial Teacher Training, curriculum design and academic research, identified two separate key purposes of routine OTL: CPD and QA. The QA function centred on the appraisal of “teaching competence” (p97) most often based on “performance criteria” and linked to a “hierarchy of levels” (p98). QA observations tended to be based on a “supervisor–subordinate model” (p98) and could be linked to “disciplinary action against teachers judged to be incompetent” (p100). In contrast, the CPD function concentrated on “fomenting critical analysis of [the teacher’s own] professional practice”; in this model, practitioners might be observed by managers or by their peers.

Whilst preserving Wragg’s key distinctions, Gosling (2002, p4-5) suggested a more nuanced trio of models of what he termed “Peer Observation”: evaluation, development and peer review models. In general terms, Gosling’s peer review model might be most closely identified with Wragg’s CPD function: consisting of “Non-judgemental, constructive feedback” and focussed on an “engagement in discussion about teaching” and “self and mutual reflection” (ibid). The evaluation and development models both feature considerable external judgement and have significant consequences for the observee, with a “pass/fail” (ibid) result: these might both be seen to constitute Wragg’s QA function, divided here to distinguish between Gosling’s differentiated purposes of OTL. In practice, within the setting of Newbold, the evaluation and development models of OTL were carried out simultaneously, except in the very rare circumstances of the later stages of a competency procedure.

O’Leary (2012) criticises Gosling’s use of the term ‘peer’ in describing these models, because “only his third model ... can be described as a genuine example of peer observation” (p802-803) in terms of commonly held understanding. However, I believe it should be noted that in the cases under consideration here, most cases of OTL were undertaken by ‘peers’, at least in a qualified sense. All those carrying out observations at Newbold were also classroom practitioners who were also observed in turn. Observers were seen to have numerous roles: they were fellow teachers, senior colleagues and part of the management hierarchy. This conflation of roles, and models of approach, might also be seen as contributing to the perceived thwarting of the efficacy of OTL within Newbold to deliver the peer review model/CPD function. In both Gosling and Wragg, this function was seen to be better carried out in an approach free from QA judgmentalism.

It is difficult to give “non-judgemental, constructive feedback” when you are also giving a grade and making “pass/fail” decisions (Gosling, 2002, p4-5). The positive effects of feedback have been shown to be largely negated by over-emphasis on grading, which “may encourage an emphasis on quantitative aspects of learning, depress creativity, foster fear of failure, and undermine interest” in its subjects (Butler and Nisan, 1986, p215); most people perform better when the focus is on “constructive, specific information about competence” rather than on summative grades (ibid).

## **2.4 Participants’ experience of OTL**

My primary interest was in how observers made judgements about the teaching they see in classrooms. There was little directly about this in the literature with most studies focussed on the effects of being observed, rather than the process of observing. These

works were considered relevant because all the observers in this study have been and continue to be classroom practitioners as well as observers and it seemed reasonable to assume that their judgements and approaches to observation will have been informed by their own experiences as practitioners who are observed and their concerns for the tutors they manage.

#### **2.4.1 Observers' experience of OTL**

Cockburn (2005) points out that practitioners predominantly associate OTL with QA and the question of competency and that, therefore, "it is not surprising anxiety is commonly associated with the process" (p374). He goes on to outline a series of concerns identified by his practitioner interviewees, including increased "scrutiny", and the "artificiality" of the OTL process (p377). Practitioners viewed OTL "as a bureaucratic exercise with little genuine concern for teacher development" and therefore, at best, "a 'necessary evil'". Practitioners derive positive benefit from OTL when "it throws light on activities central to the professional role" (ibid, p382), when the process became a "Reflective Mirror" (ibid, p381). However, for these benefits to be felt, the judgmentalism inherent in the QA purpose of OTL would have to be reduced. Although Cockburn does not specify how many interviews he carried out, referring only to a "range of professionals deeply involved in the classroom observation procedure" (p374), the interview quotes he provides are compelling and inclusion is therefore justified on qualitative grounds.

O'Leary's (2013a) examination of graded OTL in 10 FE colleges also focusses on the effect of the process on practitioners. O'Leary found that the QA purpose of OTL was the most commonly recognised by the vast majority of practitioners; although there was

also “an expectation that observers will provide formative feedback” (ibid, p703-4) – a combination that many observers found problematic (p707). O’Leary states that practitioners found the experience “divisive” (p710), “difficult to manage” (p707), and found evidence of what Jeffrey and Woods (1998) referred to as ‘playing the game’ (p709).

Likewise, the University and College Union’s (UCU) wide-ranging survey of FE practitioners (O’Leary, 2013b) found that graded OTL, carried out by managers, was overwhelmingly the most common model, in 77% of instances (p41). This survey confirmed the largely negative attitudes of observees to the experience of graded OTL mentioned above, with 67% of all respondents, which included observers and small numbers of senior managers, agreeing that graded OTL “should no longer be used as a form of teacher assessment” (p45).

Other research has not been so negative and O’Leary’s results may be the result of asking practitioners to give their ‘feelings’ about an experience that was, by its nature, intrusive on their day-to-day practice and stressful in terms of its high-stakes consequences. Cockburn (2005), for example, found that “positive responses outweighed the negative by 35%”, suggesting that many, if not most practitioners valued some aspects of the process, if not the totality. UCU’s more recent survey found far less enthusiasm among practitioners, but a still substantial 40% of respondents supported the contention that graded OTL was “essential for monitoring the quality of teaching and learning” (O’Leary, 2013b, p45). However, as O’Leary contends, this perhaps higher than expected support for graded OTL may be due to familiarisation:

“as graded observations have become normalised in FE in recent years, so staff have become increasingly conditioned to expect to be graded on their classroom performance” (O’Leary, 2013b, p43)



#### **2.4.2 Observers' beliefs about OTL**

There appears to be a lack of research within the literature specifically focusing on the observer experience. Where observers have been interviewed, or the process of observing featured, the focus of the studies themselves has usually precluded or minimalised a consideration of how or what observers feel about their OTL decisions.

Boocock's (2013) study, in common with my own, focussed on the OTL practices within one, albeit large, FE institution. A range of subjects were interviewed, from classroom practitioners to senior management, and liberal use made of verbatim quotes. However, the emphasis throughout is in line with the researcher's critical theorist agenda, on examining the political "neo-Fordian" purposes presumed to be behind the OTL process. Consequently, Boocock's study tended not to ask observers about why and how they reached their decisions, although there were some telling moments. For example, Boocock demonstrated that observers were aware of their observees' tendency to 'play the game' in their observations: "it is a show that you put on for when someone comes in and that is all it is really and if you know how to play the game you will be fine" as one observer put it (Boocock, 2013, p490). There was also a feeling amongst observers that "grading dominates" the process, to the detriment of the CPD function, which was "de-professionalising, controlling and not trusting of staff with what they do" (ibid, p 493). For this reason, it was common for observers to believe that "a less threatening OTL system, such as peer-observation, might be more effective" (ibid, p495).

O'Leary's (2013) more substantial study also included some interviews with observers. However, his focus was on the effects of OTL policy on practitioners rather than the OTL decision-making process itself. However, the study does confirm that there was a tension experienced by observers between the QA and CPD functions of OTL: one

observer stating that it was “difficult to manage two very different roles for observations under the one umbrella” (O’Leary, 2013, p707); and, in a rare comment upon the feelings of an observer within the literature: “I’m not sure I succeed in achieving both and I feel guilty about that” (ibid). O’Leary found that most observers expressed a desire to prioritise the CPD function of OTL but that “their ability to uphold such a commitment was compromised by the prioritisation of the QA agenda in colleges, coupled with practical time constraints” (ibid).

#### **2.4.3 The relationship between observer and observee**

There was little in the literature directly related to the potential effect of the relationship between the observer and the observee in an in-house OTL context. The creation of a good working relationship was often cited as beneficial, but the problem of observers making judgements of observees they might have known well for years was not dealt with. In the context of the less high-stakes mentoring relationship, Cullimore and Simmons (2010) found that mentors expressed anxiety, and even guilt, “about the need to be critical of their colleagues” and that “relationships in the staffroom can get in the way” (p231). The mentors were even more anxious about their judgements when these would have to become part of official feedback, one mentor asking whether, in the case of picking up on ‘problems’, they should, “pass them because this person is a colleague and friend, or fail them because you know they’re not ready?” (ibid, p232). It would appear reasonable to assume that given the high-stakes involved, and the status differential between observer and observee, that the effects may be even more significant in official graded OTL.

## 2.5 OTL decision-making

This section relates to RQ 2: How do observers explain their OTL decision-making? It provides a discussion of theory from the literature illuminating the area of OTL decision-making. The concepts of ‘criteria’ and ‘expert judgement’ are discussed as key terms in my understanding of decision-making as it applies to OTL.

Despite refining and adapting search terms, I was unable to find any study focussing specifically on how and why OTL decisions are made. There are countless guides as to what to look at and consider when observing in the classroom; but nothing on how observers actually come to their decisions. I address my second RQ to exploring this apparent gap in knowledge.

### **The problem of observation: “what are we looking at?”**

This is, perhaps, a curious question. Surely, observers are looking at ‘teaching and learning’? The visiting observer can certainly *watch* the behaviour of the teacher and the learners, but does this amount to the same thing as *seeing* “teaching and learning”?

This question is much debated in the literature. Coe (2014) argued that “learning is invisible” and could only be judged by “teacher behaviours” and “observable proxies” of arguable provenance. According to O’Leary (2006) observations of teaching concentrated on observable “low inference or low order factors”, such as seating arrangements, which “tell us very little, if anything at all, about teacher behaviour and the learning process itself” (p195).

These writers critique OTL and question its usefulness as a measure of ‘effective teaching’. However, majority opinion throughout the literature, as well as the world of practice, has been that it is possible to observe teaching and to draw useful conclusions

from doing so. Shulman (1988) pointed to this usefulness, stating that "in principle, classroom observations can reflect the full complexity of teaching" (p19). However, in order to be able to do this, the system of OTL must "first and foremost be faithful to teaching" (p17); by which he meant based on "watching real teaching in real classrooms" (p19).

### **Models of decision-making**

In psychology, two models of decision-making are often presumed to function:

- Heuristic judgements based on "lists of attributes we believe are defining and characteristic" (Hastie and Dawes, 2010, p106), and might here equate with the use of judgement-criteria.
- Holistic judgements which "draw intuitively on an agent's experience" (Provis, 2010, p3), and represent a more personal and immediate response to the observed session.

It is a matter of philosophical perspective whether one believes that heuristic, criteria-based approaches are a qualitative advance on the holistic; or, in contrast, whether the more intuitive decision-making which some believe characterises expert decision-making was in fact an evolution from the heuristic. Clearly there are claims to be made for both viewpoints: it is undoubtedly more advanced to base one's decisions on an understanding of the realities of a given situation, rather than making uninformed random judgements based on nothing. However, this distinction is true only at the most basic level; in the context of professional decision-making, a sophisticated level of judgement may be accomplished with apparent automaticity, with the criteria of

decision-making being perhaps largely internalised and accessed unconsciously by the experienced observer.

Hastie and Dawes' (2010, p4) provided the analogy of driving a car: learner-drivers consciously attend to every detail – to every criterion of 'effective driving'; with experience, the expert-driver develops "automatic thinking": an unconscious reliance on accrued experience and know-how. The holistic then, when applied by an experienced practitioner, is the more advanced form of judgement *in a practical situation*. However, Hastie and Dawes stressed that "any significant intellectual achievement is a mixture of both automatic and controlled thought processes" (*ibid*, p5).

### **2.5.1 Criteria**

The influential early American educationalist, John Dewey, observed that evaluative judgements "spring from the immediate and inexplicable reaction of vital impulse and from the irrational part of our nature", that they are basically "a-rational" (1939, p18). Acknowledging this, Royce Sadler (1985) suggested that "recognition" based on this immediate and impulsive form of judgement, "is the primary evaluative act and predates any criteria": "things can often be recognized as excellent or beautiful before the rules for excellence or beauty are formulated explicitly" (p291).

Royce Sadler's argument was that the formulation of these "rules", or criteria, allow judgement about such abstractions to proceed in an apparently more rational way.

Criteria arise initially from "the process of reflecting about and providing rationalizations" (*ibid*, p291) for the impulsive holistic evaluations we make; the "criteria are initially descriptive", but in time they "begin to function normatively" (*ibid*). In this

way, the use of criteria allows decision-making to evolve “from reactive or holistic valuations to rational valuations” (ibid, p293).

Heuristic decision-making necessarily involves a reliance on a set of consciously created criteria: “the lists of attributes that we believe are defining and characteristic” (Hastie and Dawes, 2010, p106) of any given concept, such as ‘effective teaching’. Hastie and Dawes acknowledged that “heuristics are efficient” (ibid, p88) and “useful” (ibid, 106), in that they offer straightforward demonstrable evidence for judgements made, but also that this kind of thinking was “partly fictional simplification” (ibid) because a mixture of the holistic and heuristic was always really involved, based on cognitive tools that were “acquired over a lifetime of experience” (ibid, p88).

The advantage of using criteria is clear, especially when the judgement in question is to be official and carry a claim of authority, as in the case of high-stakes OTL decisions. The appeal to criteria affords the appearance of quasi-scientific reliability. Ultimately, however, Sadler concluded that because “criteria are mental constructions which cannot be experienced or enjoyed directly, it is therefore necessary to consider the reasonableness of demands that all evaluations be rational” (ibid, p294).

As has been mentioned, the conduct of Ofsted inspections has been an influential, and arguably dominant, influence on the practice of OTL in many institutions. For Ofsted, in an early guide for inspectors, criteria “amount to standards for good practice” and in “reaching overall judgements, all the relevant criteria should be considered” (OFSTED, 1995, p45). This would appear to locate the decision-making process squarely in the heuristic approach. However, the 1995 guide went on to say that “the criteria can be used in different ways”; they could indeed be used “cumulatively to build up an overall judgement” but alternatively they could also be used “to test and check” judgements already made, presumably on the basis of a more holistic approach (ibid). In more

recent versions of the guidance, this equivocation disappeared and the claim to the heuristic approach has become privileged, but there was still an allowance that the criteria alone cannot form the sole basis of judgement: the criteria “are not a checklist” (OFSTED, 2015, p28) and it was necessary that “inspectors will exercise their professional judgement” (ibid, p3).

Wilcox and Gray’s (1996) interviews with serving Ofsted inspectors revealed something of the reality of their decision-making process: “I have a crib sheet, we all have a crib sheet”, said one in reference to the criteria, “but after you have done a lot, you know what you are looking at and what [grade] you’d be giving them” (p73). This suggests that familiarity with the process of carrying out graded observations leads from the heuristic criteria-focused approach towards the more holistic “automatic thinking” style associated with greater experience. Another inspector summed it up by admitting: “I think in the end inspectors make totally subjective judgements” (ibid), an admission consistent with Ofsted’s appeal to professional judgement. As Nixon and Rudduck (1992) pointed out, there appeared to be: “a strong tacit component to professional judgment that cannot be circumscribed by the simple expedient of making explicit the criteria by which it is supposed to operate” (p135). The implication being that Ofsted grade-decisions, and the OTL decisions that attempt to mirror them, cannot be free of “the unpredictability and diversity of professional judgment” (ibid) just because a set of criteria is put forward as evidence or justification.

The validity of a criteria-focussed approach, specifically Ofsted’s, was subjected to a ‘Wittgensteinian Critique’ by Gilroy and Wilcox (1997). They suggest that Ofsted’s claim to validity, based on “consistent application of the criteria” (p27), was undermined by the fact that “such criteria are tacit” and when we try to explain and justify them “there is a natural tendency to use other criteria to explain the meaning of

the criteria being questioned” (ibid, p29). This process leads to logical absurdity, in which criteria “would, in turn, require further criteria” in order to explain them, leading ultimately to “an infinite regress of criterial explanations” (ibid, p29-30). The authors’ findings that in practice experienced observers developed “‘rules of thumb’ to deal with these issues” (ibid,p28), might reasonably be thought to indicate that, despite the appearance of heuristic, ‘objective’, criteria-based decision-making, much of how inspection judgements were made was, in fact, located in the holistic, experience-based approach of the individual observers.

Strong et al. (2011) provided one of the only studies to consider the criteria used by observers in their OTL decision-making. This study was not specifically looking at the decision-making process of experienced observers, and was not based on data taken from practice. Strong et al. set up a series of quasi-experimental situations which compared the results from the observation of video clips of teaching practice. In their first experiment, a total of 100 “judges” took part, including some practitioner-observers, but also school administrators, education professors, parents, trainee teachers and “adults with no formal connection to education” (p371). **Table 2.1**, below, shows Strong et al.’s list of “teaching strategies commonly cited as influencing judgments” arising from this research (in diminishing order of frequency):



<b>Teaching strategies commonly cited as influencing judgments:</b>
Accesses students' prior knowledge
Creates stimulating classroom environment
Has active interaction with students
Moves around classroom
Enables students to generate ideas
Uses visuals and manipulatives
Checks for student understanding
Has clear objectives
Presents concepts clearly
Exhibits equity
Differentiates instruction

Table 2.1: 'Teaching strategies commonly cited as influencing judgments' (from Strong et al., 2011, p374)

The research design of this study, including “judges” other than observers “to determine if their relationship to education affected their judgments”, was quite different from my own, which was based on data arising from practice in the field. However, this list does feature elements of ‘effective teaching’ commonly found in the literature (see Section 2.7, p49).

### 2.5.2 Expert judgement

As the experience of any practitioner increases, his or her judgement processes appears to evolve into what Hastie and Dawes (2010) termed “Automatic Thinking”: a way of operating at a high level of expertise in which the “thought processes are so automatic that we are usually unaware of them” (p4). Sadler (1985) stated that it was “possible for reliable judgments to be made even when no criteria are used” such judgements could be considered “valid to the extent that the evaluator is accepted as authoritative and competent” (p286).

In their seminal 1980 study, Dreyfus and Dreyfus described the process by which practitioners, in any particular field, were able to function in their roles as they accrued experience and expertise. From the situation of the novice - who must consciously follow explicit rules and guidelines - practitioners reached the final stage of development, "Expertise". The expert practitioner was one whose:

"repertoire of experienced situations is so vast that normally each specific situation immediately dictates an intuitively appropriate action."

(Dreyfus and Dreyfus, 1980, p12)

Brown and McIntyre (1993), recognised that the intuitive nature of expert knowledge presented some difficulty in trying to understand the processes involved:

"While we recognise that there are those with mastery of some aspects of teaching, we have no coherent account of what they are masters of or how they achieve what they achieve."

(Brown and McIntyre, 1993, p13)

This observation applied specifically to the process of teaching, but might reasonably be applied to the process of observers trying to understand what is going on in a classroom and attempting to come to an OTL judgement.

More recently, Ainley and Luntley (2007) have elaborated this observation into what they have called "attention-dependent knowledge", a largely invisible phenomenon which practitioners themselves found difficult to explain: it "not only is *not* reflected in what is written down in lesson plans, but cannot be written down" (ibid, p1127). As Yeager (2000) has suggested, "experienced teachers simply possess a great deal of knowledge and understandings that they have not formally articulated" (p352). Since

Ainley and Luntley stated that this kind of knowledge “can only be revealed in classroom practice” (Ibid, p1128), it seemed reasonable to conclude that it could only be assessed through the observation of that practice.

In contrast, Strong et al. (2011) provided an important critique of the reliance on expert judgement, pointing to their own, and others’, findings of weak correlations between OTL judgements and ‘effective teaching’ as indicated by learner achievement, and suggested that well known psychological phenomena might be applied to help explain this apparent gap, specifically:

Confirmation bias	A tendency to seek, embellish, and emphasize experiences that support rather than challenge already held beliefs
Motivated reasoning	We look more sceptically at data that do not fit our beliefs than those that do
Inattentional blindness	In which people fail to notice stimuli appearing in front of their eyes when they are preoccupied with an attentionally demanding task

*Table 2.2: “Cognitive operations that influence judgements of human behaviour” (adapted from Strong et al. (2011, p369))*

In other words, on occasion, observers might see what they expected to see, or fail to see what did not fit with their expectations; and that they might use limited evidence to confirm what they thought they already knew about their observee’s abilities. In terms of decision-making and judgement, it is worth remembering that being human, observers are prone to such biases.

## 2.6 Wisdom of Practice

“The road we drive on has bends and potholes not included on the map.”

(Schwab, 1971, p496)

Here, Schwab provides a beautiful metaphor as much for life, as for the classroom experience. It points towards the truth that no map, no set of rules, can possibly capture the completeness of a complex human experience like teaching. It may even be true that no such abstraction will ever effectively focus on what is really important in such a context.

This is an appeal to the existence of what the anti-positivist polymath Michael Polanyi called “tacit knowledge”; that “which everyone uses in the ordinary course of living” (1962, p295); and which some theorists (Kinsella and Pitman, Higgs, both 2012) have linked to the Aristotelian concept of ‘phronesis’; an intellectual virtue which:

“involves deliberation that is based on values, concerned with practical judgement and informed by reflection. It is pragmatic, variable, context-dependent, and oriented toward action.”

(Kinsella and Pitman, 2012, p2)

For the expert practitioner, this phenomenon becomes “craft knowledge” (Leinhardt, 1990, p18): “the wealth of teaching information that very skilled practitioners have about their own practice” including “deep, sensitive, location-specific knowledge of teaching” (ibid). Higgs (2012) saw such ‘practical wisdom’ as “the ineluctable nexus between practice, judgement, and knowledge” (p8).

Lunenberg and Korthagen (2009), in their theoretical analysis, also emphasised the link between ‘practical wisdom’, ‘theory’ and ‘experience’. ‘Practical wisdom’ was defined as

a “sensitivity for and awareness of the essentials of a particular practice situation” that was “intrinsically connected to specific phenomena occurring in the here-and-now”; they made the point that this type of intuitive understanding was hard to make “fully explicit” (p227). ‘Theory’ here is taken to mean knowledge that “involves logical structuring, such as the formulation of definitions and logically derived propositions”, related to “insights developed by others”, perhaps accessed by practitioners in teacher training and CPD sessions (ibid). Finally, ‘experience’ was defined as that which was gained from “operating in the real world, in practice” and encompassing “both the environment (e.g., the classroom) and one’s own inner reality while relating to this environment” (p228).

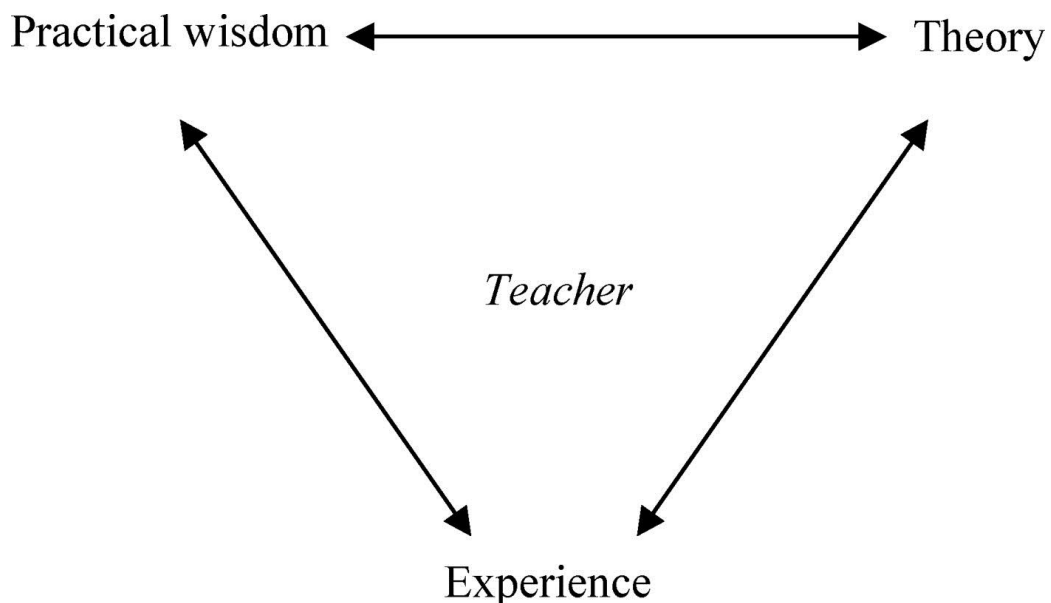
Lunenberg and Korthagen’s ‘practical wisdom’, ‘theory’ and ‘experience’ can be usefully compared to what Shulman (1987) termed the “categories of the knowledge base” (p8) of practitioners:

<b>Categories of the knowledge base</b> (Shulman, 1987, p8)	<b>Experience, theory, and practical wisdom in teaching</b> (Lunenberg and Korthagen, 2009, p227-8)
Content knowledge	Theory
Curriculum knowledge	
General pedagogical knowledge	Practical wisdom
Pedagogical contents knowledge	
Knowledge of learners and their characteristics	Experience
Knowledge of educational context	
Knowledge of educational ends, purpose, and value and their philosophical and historical grounds	

*Table 2.3: Comparison of Shulman’s ‘Categories of the knowledge base’ (1987) and Lunenberg and Korthagen’s ‘Experience, theory, and practical wisdom in teaching’ (2009)*

It can be seen that Shulman’s conception of knowledge/experience, split as it is across three distinct categories, was a good deal more practical in nature than Lunenberg and Korthagen’s, and distinguished clearly between generalised understanding and a practitioner’s knowledge of particular settings, situations and people.

Lunenberg and Korthagen suggested a triangular relationship (See **Fig. 2.4**, below) between their three elements in which it was the practitioner “who makes them all come together (or not) in the day to day practice” in the classroom (2009, p229).



*Fig. 2.4: The triangular relationship between practical wisdom, theory, and experience. (Lunenberg and Korthagen, 2009, p229, fig.1)*

Intrinsic to Lunenberg and Korthagen’s model is that this was a “complicated, non-linear process, in which context, i.e, workplace conditions, play an important supporting or inhibiting role” (ibid).

Brown et al. (1988) also dealt with the subject of what they called “professional craft knowledge”, defined as “the ordinary things which experienced teachers do spontaneously in their classrooms ... and how they conceptualise their own classroom teaching” (p3). Buchmann, (1987) talked about “the folkways of teaching” which “describe 'teaching as usual', learned and practised in the half-conscious way in which people go about their everyday lives” (p151). For convenience, I shall refer to this idea from now on by the nomenclature by which it is, these days, most often referred to: ‘Wisdom of Practice’.

According to Shulman (2004, p505), the phrase, ‘Wisdom of Practice’ was coined by the academic and philosopher of science, David Hawkins. Hawkins (1966) pointed out that “the personal knowledge of practitioners was [often] significantly deeper than anything embedded in the beliefs and writings of the academically learned.” (p3). He believed that ‘effective teaching’ “owe[d] little to modern theories of learning and cognition and much to apprenticeship, on the job inquiry, discussion, [and] trial ...within a common-sense psychological framework” (p4). It is this idea which is at the heart of the concept of Wisdom of Practice.

Throughout his career, educational psychologist, Lee Shulman has built upon and elaborated the idea of Wisdom of Practice, which he defined as "a source for understanding the complexities of skilled performance" (Shulman (a), 1987, p257). Shulman drew on observations and research in a range of fields. In medicine, for example, he found that "problem formulation and hypothesis generation" were "closely related to the physician's substantive knowledge base and specific experiences in a particular domain" (ibid, p256).

Shulman found that teachers often ignored or adapted theoretical good practice in the classroom. For example, experienced practitioners may find the practice of lengthened

question-wait times unattractive because “they bring with them an increase in the problems of classroom behaviour” (ibid, p263), this could be seen as an experience-based, pragmatic adaptation to real-life classroom factors. As Weimer (2001) has stated, “most of what is learned about teaching occurs as on-the-job training, in the isolation of the individual’s classroom, alone and without supervision” (p46), although in my experience some practitioners learn more effectively from this than others.

In Shulman’s view, Wisdom of Practice did not present “some uniform, monolithic image of ‘good practice’ ... wise practitioners vary” (ibid, p265). Instead the Wisdom of Practice approach is an attempt “to understand the grounds on which [good practice] rest[s].” This has implications for OTL; Shulman stated that “any system of teacher assessment, however reliable, economic, or efficient, must first and foremost be faithful to teaching” (1988, p340). He believed that “in principle, classroom observations can reflect the full complexity of teaching, but they barely achieve their potential in practice” (ibid, p341). The strength of OTL is that it is based on “watching real teaching in real classrooms directly” (ibid, p342), but its value is dependent on Shulman’s recognition that teaching could not be reduced to a set of criteria on a check list: “teaching is more than classroom management and organisation; more than knowledge of subject matter”(ibid, p347). For Shulman, teaching is an altogether more complex activity:

"Teaching is impossible. If we simply add together all that is expected of a typical teacher and take note of the circumstances under which those activities are to be carried out, the sum makes greater demands that any individual can possibly fulfil. Yet teachers teach... "

(Shulman, 1983, p153)



It is this "role complexity" (ibid, p154) that makes teaching "impossible" in Shulman's terms, or rather makes it impossible to unpick all the elements that comprise the activity at any point. To understand what constitutes 'effective teaching', it is necessary to ask: "what is the size of the irreducible kernel of professional judgement without which teachers will not be able to respond adequately to the unpredictable complexities of life in the classroom?" (ibid, p159). This "irreducible kernel" is the Wisdom of Practice.

Whilst it might be tempting to romanticise the concept of Wisdom of Practice, it must be remembered that the concept of Wisdom of Practice is far from uncritically accepted. Leinhardt (1990) warned that Wisdom of Practice might include much which is "fragmentary, superstitious, and often inaccurate" (p18). Weimer (2001) also offered a critique suggesting that Wisdom of Practice "generally ignores the peculiarities of fit"; in that "individual practitioners seldom [have] any sense of why some strategies, policies, practices, ideas, techniques, and approaches work in some contexts and not in others" (p49); and because "poor and inadequate assessment techniques and approaches" meant that practitioners "tend to rely first and foremost on their own very personal assessments of how well things worked [making] that assessment while ignoring their personal investment in teaching, which jeopardizes their ability to be objective" (ibid, p52). Weimer's objections point to the surely reasonable conclusion that the Wisdom of Practice is a far from objectively experienced phenomenon, which might resist direct enquiry and attempts to subject its content to revision and evaluation. This is not, however, to say that the concept has no useful application or that it does not function as an untidy yet powerful metaphor for the subjective experience of many practitioners.

## **The nature of Wisdom of Practice *in situ***

To summarise, the literature on expert knowledge and the Wisdom of Practice appears to characterise the phenomenon in the following general terms:

- It is spontaneous (Brown et al., 1988)
- It is “automatic thinking” (Hastie and Dawes, 2010)
- It arises from practical experience (Shulman, 1987)
- It is intuitive (Dreyfus and Dreyfus, 1980)
- It is unconscious (Buchmann, 1987)
- Practitioners find it hard to explain (Yeager, 2000)

All of which suggests that Wisdom of Practice is experienced by practitioners on a level which they find difficult to consciously access and, therefore, explain; making it difficult for a researcher to directly address through questioning. However, it may be possible to apply the characteristics referred to above, in order to investigate whether such a process appears to function in OTL decision-making.

## **2.7 ‘Effective Teaching’**

This section relates to RQ 3: What elements of ‘effective teaching’ do observers focus on in evidencing OTL decisions?

Whilst, there was a great deal of literature on what constitutes the elements of ‘effective teaching’, I have found little research that examines the extent to which these elements figure in actual observers’ decision-making. I will therefore outline and discuss the range of literature on the subject, and attempt to distinguish the key elements of

‘effective teaching’ arising throughout, which may be usefully compared with the elements identified by the observers in my research.

Throughout this thesis I have put the words ‘effective teaching’ within inverted commas. I made the decision to do this because I have not wished to presume that there exists a definitive grasp on this concept. Part of the purpose of this research was to find out what features of teaching observers were using as the bases of their OTL judgements – to unambiguously assume that these elements are those that constitute ‘*effective teaching*’, would appear to beg the question. Instead, it was my intention to discover what criteria the observers in the research used to evidence their judgements, and to compare these to those elements identified throughout the literature on this subject. Whilst not offering an objective ‘proof’, a degree of consensus across the range of literature, and among the observers in this study, might be presumed to offer an indication that what was being referred to, indeed, equated to ‘effective teaching’.

### **Elements of ‘effective teaching’ in the Literature**

As part of my review of the relevant literature, I considered a wide range of academic studies, official documentation and professional guidance literature – all dedicated to the task of defining the parameters of ‘effective teaching’, and to identifying the elements by which the phenomenon of ‘effective teaching’ might be recognised and/or assessed. A full list of the sources considered is included in **Appendix 5** (p222).

Whilst none of the sources cited need be considered definitive or beyond criticism either in terms of their methods or their conclusions, these sources might reasonably be thought to represent an indicative sample of the most highly influential opinions and findings within the literature on ‘effective teaching’.

The sources included fell into three distinct but not unrelated groupings:

**Governmental/Agency guidelines and advisory documentation:** Sources such as Ofsted (2014): foundational materials issued by ‘authority’ bodies exerting considerable influence on institutions, forming the basis of management targets and setting the agenda for inspection training and priorities within OTL regimens.

**Professional guidance literature:** published mainly in the form of books or pamphlets, and aimed at practitioners, trainee teachers and other interested professionals. An excellent example is Dunne and Wragg (1994) which, despite its age, still provides a wide-ranging and authoritative account of the practical skills required of classroom practitioners. Also included here are reports intended for the more specialist consideration of administrators and policy makers, such as Coe et al. (2014) which presents an exhaustive overview of the field and puts forward several recommendations that are referred to elsewhere.

The North American perspective is represented by Teddlie et al. (2006), Danielson, C (2011) and the CLASS project (Pianta et al., 2008). Danielson and CLASS are both reports connected to the foundation of formal structured observation techniques, which have been widely adopted in schools across the USA. Regardless of the debate as to the efficacy of this approach to OTL, both systems have been based on extensive research and reviews of existing studies in the field.

**Academic studies:** the oldest academic source included is Shulman (1987); significant within the field and central to the concept of Wisdom of Practice.

The seminal nature of Shulman’s work means that it has continued to be

extremely relevant. Brown and McIntyre (1993) has been included as part of a body of literature that has refined and expanded Shulman's work.

Black and Wiliam's (1998) study is the foundational text of the Assessment for Learning (AfL) movement in the UK. This work was based on an extensive and exhaustive systematic review of a wide range of studies, including into the field of 'effective teaching'. Their recommendations, forming the basis of the AfL approach, are included as they became highly influential in government policy, Ofsted guidelines and practice in schools throughout the 2000s and has been adapted and adopted widely within the FE sector.

A very different approach, but nonetheless influential, has been Hattie's (1999) statistical analysis of the impact of various teaching strategies and behaviours. His list of the most effective approaches has been included both because of the claims to validity made within Hattie's methodology, and on the basis of the wide-ranging influence of this work which is widely cited throughout the literature.

As I have stated, there has been considerable cross-fertilisation within and between sources in the different categories. Wragg's work, for example, has been widely influential both within the academic field and as the basis of official advice on classroom practice. Several of the academic studies contain references to one or more other academic source also cited in this review - Black and Wiliam's (1998) work, for example, is based on a wide-ranging review of other studies – and this could be seen as a weakness here, potentially leading to 'double counting' some criteria. It would have been surprising if, within such a relatively contained field, there were not considerable overlap between sources – and indeed the homogeneity of the criteria of 'effective teaching' contained in the sources is, in itself, indication of the level of agreement in the

literature. The same criteria occur again and again, the same advice and guidance to practitioners, the same dos and don'ts. That a high degree of agreement was found between the aspects of 'effective teaching' focussed on in these reports and those of the rest of the cited sources goes some way to suggesting that there is a wide sharing of the definitions of 'effective teaching'.

Pragmatically, I found it necessary to include sources that were primarily focussed on the school experience because the majority of the work that has been done on 'effective teaching' has been located in that sector, for the obvious reason that this is where most teaching takes place and most of the money spent on education goes. Excluding school-focussed research would have meant excluding most of the key texts in the field. The most obvious practical effect of this inclusion was that some of the elements cited in the literature were not largely applicable to or comparable with my adult education setting: specifically pedagogical and pastoral concerns and what Danielson (2007) calls "managing student behaviour".

These exceptions aside, my personal belief was that most elements of 'effective teaching' can be considered universal across the range of sectors. Skills such as 'classroom management', 'planning' and 'questioning technique' are indicative of 'effective teaching' no matter who is being taught or where. This belief is based on my experiences of teaching in both schools and in FE, and is supported by the guidance provided by the Education and Training Foundation (ETF), that there is "a set of universal standards that are applicable to all teachers and trainers in any part of the Education and Training sector" (ETF, 2014, p6).

### Most frequently occurring elements of ‘effective teaching’ in the literature

From this survey of the literature on ‘effective teaching’, I was able to collate a table which tallied the individual citations of each key element. **Table 2.5**, below, shows the most frequently occurring elements of ‘effective teaching’ taken from the 24 sources referred to above (see complete list in **Appendix 4**, p221).

Element	Frequency
Assessment	21
Planning	20
Feedback	19
Learning environment	19
Activity	18
Aims	18
Challenging work	18
Clear instructions	18
Engagement	18
Subject content	18
Differentiation	17

*Table 2.5: most frequently occurring elements of ‘effective teaching’ in the literature*

The degree to which occurrences of the most frequently appearing elements were ‘bunched’ together was remarkable, in my opinion. This apparent consistency across the 24 sources as to the key elements of ‘effective teaching’ suggested to me a level of agreement within the literature that might constitute support for the existence of a Wisdom of Practice within the field of practice. I believed it would be instructive to compare this list of elements with the findings of my research.

## Explanation of some key elements

Although most of the elements of ‘effective teaching’ referred to throughout, and comprising the judgement-criteria used by the observers in this study, will be instantly recognisable to anyone with a passing familiarity with education, it might be useful to briefly explain some of the key terms:

Criteria	Comments
<b>Aims</b>	“Statements which encapsulate the educational value and worth of lessons” (Hickman , 2009, p7). In practice, observers are looking for aims and objectives set for lessons and for individual learners. Sometimes this is about sharing the aims with the class.
<b>Assessment</b>	“The process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning” (ARG, 2002, p1). Observers might comment on any assessment activity or recording.
<b>Differentiation</b>	“The organisation of teaching programmes and methods specifically to suit the ability and aptitudes of individual learners” (Hickman, 2009, p29). Usually focused on materials and activities planned to reflect different levels of ability.
<b>Feedback</b>	“Process through which students learn how well they are achieving what they need to do to improve the performance” (Isaacs et al., 2013, p61). Can refer to verbal or written comments by teacher.
<b>Learner voice</b>	This refers to learner views gathered as part of a routine observation, which will include a short Q&A with learners.



<b>Reflective practice</b>	“The process of learning through, and from, experience towards gaining new insights ... the practitioner being self-aware and critically evaluating their own responses to practice” (Finlay, 2008, p1). In practice, focussed on existence and quality of learner diaries and/or teacher evaluations of lessons.
<b>Relevance</b>	Based on the widely held assumption that adult learners “are practical and need to focus on what is important to them” (Cercione, 2008, p145). Focussed on learning materials and context that connect with learners’ lives and experiences.

*Table 2.6: Explanation of some key elements of ‘effective teaching’*

## 2.8 Outcome measures of ‘effective teaching’

This section relates to RQ 4: “How do observers’ judgements relate to other potential outcome measures of ‘effective teaching?’” It explores recent developments leading to a widening focus in the assessment of ‘effective teaching’ to include measures other than graded OTL. Specifically, these potential ‘outcome measures’ are: Accreditation and Achievement, valued by Ofsted and recommended by some authorities in the Literature; and attendance rate. Both of these were managerial priorities with reputational and financial implications for many educational institutions. Another potential measure may be located in the collation of learner satisfaction ratings; however, in practical terms, because no systematically collected data for learner satisfaction exists in Newbold, I have excluded this.

Although the question of alternative measures of ‘effective teaching’ has been covered widely in recent literature, there does not appear to have yet been a study exploring data based on actual practice, as opposed to quasi-experimental or statistical approaches.

## **Accreditation and Achievement and ‘Effective Teaching’**

Recently, there has been a movement away from the reliance on graded OTL as the principal measure of ‘effective teaching’, at least on its own. Crucially, Ofsted have stated that inspections “will [no longer] grade the quality of teaching”, with classroom observation “supplemented by a range of other evidence” (Ofsted, 2015, p45). Coe et al. (2014, p3), in their exhaustive and wide-ranging review of the research underpinning the field of QA in teaching, conclude that OTL judgements “need to be used with considerable caution”, suggesting an approach based on “multiple measures” (ibid), specifically including achievement and accreditation outcomes.

Strong et al. (2011), in their elaborate quasi-experimental study of the efficacy of classroom observation, find that “there is not much evidence to suggest a strong relationship between these observation-based teacher evaluation ratings and student academic outcomes” (p368). In fact, they concluded that “in every case” the observers in their survey were “absolutely inaccurate” (p378), leading them “to question whether educators can identify effective teachers when they see them”. This was a striking finding, but it is dependent on the definition of ‘effective teaching’ used. Importantly, Strong et al. based their assumptions on their belief that achievement outcomes represented some kind of ‘true value’ of ‘effective teaching’; and this is far from obvious – rather, it represents a particular definition of ‘effective teaching’.

Fenstermacher and Richardson (2005) agreed with Coe et al. that ‘effective teaching’ should be understood as “teaching that produces learning” (p5). However, they identified two parts to the concept of ‘effective teaching’, and only one of these was dependent on demonstrating achievement outcomes. Fenstermacher and Richardson differentiated between what they called “successful teaching”, which delivered the

intended learning and results in accreditation and achievement; and “good teaching” which is “teaching that comports with morally defensible and rationally sound principles of instructional practice” (p6). They are supported in this differentiation by Berliner (2005) who also suggested that “quality teaching” consists of “two conceptually separate parts” (p207).

‘Effective teaching’ might be thought of as teaching which is both “good” and “successful”, if these terms are to be accepted. However, Fenstermacher and Richardson emphasise the separate nature of these elements: “not all instances of good teaching are successful, nor are all instances of successful teaching good” (ibid, p11). It is possible to deliver learning and achievement of results in a way that does not comport “with morally defensible and rationally sound principles of instructional practice” – arguably, perhaps, a narrow emphasis on teaching to the test. Equally, it is possible to teach in a highly proficient and technically sound manner, but for the learners to fail to show any evidence of learning because “teaching, by itself, does not produce learning” (ibid, p16).

Following Fenstermacher and Richardson (2005), it might be claimed that that which is primarily assessed in OTL, consists of their concept of “good teaching” (which is “sensitive to the learners taught, but not dependent on learning taking place” (p37)); whilst that which might be assessed in reference to accreditation and achievement, rests in their conception of “successful teaching”.

Even so, Fenstermacher and Richardson warned against a “naive conception of the relationship between teaching and learning” (p10). They proposed four critical factors for learning, all of which seem reasonable:

1. Willingness and effort by the learner
2. A social surround supportive of teaching and learning.
3. Opportunity to teach and learn
4. Good teaching

(ibid, p8)

They went on to note that only one of these factors “relates directly to the activities of the teacher” and so, “the teacher may be viewed as having a kind of limited liability for the success or failure of the learner to acquire the content taught” (ibid, p10). An over-reliance on accreditation and achievement as an indirect measure of ‘effective teaching’ might therefore be problematic. Learning requires more than just ‘successful teaching’ to take place – although, realistically, this was unlikely to be accepted in many institutions where teachers were held responsible for their learners’ results, whether this was fair or not.

Fenstermacher and Richardson’s analysis suggested that a “presumption of simple causality is more than naive, it is wrongheaded” (ibid, p10). It may be possible, therefore, to work on the principle that OTL is an assessment primarily concerned with measuring “good teaching” rather than “successful teaching”; and certainly not involved with the measurement of learning. Some factors of ‘effective teaching’ might submit to

quantitative assessment (through analysis of accreditation and achievement rates, for example), whilst others might prove more elusive and will be more readily accessible via the recognition of experienced observers through OTL.

### **Attendance rate and ‘effective teaching’**

At the time of writing, the attendance rate of learners on courses was a key funding issue within post-compulsory education, where learners can decide their own attendance. Attendance rate was considered an important element in the successful completion of accreditation, so there was an expectation that these two factors would be related. Equally, on an anecdotal level, many managers within Newbold believed that attendance of learners was directly related to the effectiveness of the teacher.

The question of what influences attendance rate is an under-researched area, presumably because the vast majority of teaching takes place in schools where attendance is mandatory and subject to legal measures to prevent non-attendance.

Most of the work that has been done in relation to non-mandatory courses has focussed on higher education settings. There are differences between the motivations and barriers affecting the attendance of university undergraduates and adult learners in the community. However, some general points might usefully be taken from the literature, as long as these differences are held in mind. It should be noted that each of the studies cited here support the premise of a link between quality of teaching and learner attendance; I could find no study that contradicted this viewpoint.

Davidovitch and Soen (2006), in a study involving nearly 10,000 completed questionnaires, albeit at only one large Israeli HE college, were able to conclude the

existence of a link between attendance rate and the perceived quality of teaching: “the higher the students' evaluation of their instructors, the higher the frequency of class attendance” (p701).

White (1992), in a much smaller study of American undergraduates, suggested that the interpersonal skills of the teacher might be key to motivating learners to attend, whilst also implying that an element of entertainment might be useful, the relationship between teacher and taught being “similar to the relationship between the actor and the audience” (p14). In support of this, Garner (2006), also focussing on undergraduates, claimed that humour had been “shown to increase attendance in class” (p178). However, whatever the benefits of pedagogic fun, White (ibid) also warns, somewhat contradictorily, that “teachers may find that being stimulating, entertaining, and otherwise creative may not ensure a high rate of attendance” (p15).

## **2.9 Summary**

This chapter has established, firstly, that there is a lack of literature directly concerning the experience of being an OTL observer, in general, and OTL decision-making, specifically. It therefore seems reasonable to assert that the findings of this study represent a contribution to knowledge.

This chapter has also given background information concerning the history and current conduct of OTL, as well as discussion of key concepts that will figure in the study later on: such as criteria, decision-making, and expert opinion. The section introducing the concept of Wisdom of Practice has begun to set it in the context of OTL, although it must be acknowledged that most of the literature available focusses on Wisdom of Practice as it applies specifically to teaching.

Finally, this chapter has established the, perhaps not unexpected, general agreement within the literature as to the key elements of 'effective teaching'. It would be interesting and illustrative to compare this set of elements to the criteria collated from the observation reports in this study and referred to by the respondents of my interviews and questionnaires. A sense of general agreement across the sources of opinion would be strongly suggestive of the existence of a shared Wisdom of Practice in operation.

## Chapter 3 – Research Design

### 3.1 Introduction

In this chapter, I outline and discuss my ontological and epistemological positions, and how these have shaped my approach in this research and my choice of methodology.

This chapter also presents details of the ethical considerations I encountered.

This chapter sets out, and discusses in turn, the following methods which were used to collect the data:

- **Single-item questionnaires:** a questionnaire consisting of one open-ended question for observer-managers.
- **Interviews:** Face-to-face interviews with six experienced observer-managers (not included in the questionnaires above).
- **Secondary data analysis** of observation reports, accreditation and achievement data, and attendance registers.

These sections address the rationale to my decision to adopt each method in reference to the literature on methodology, and describe and critique each method's application.

Tables summarising the details of the methods and of how each method was anticipated to address my research questions are included at the end of this chapter (section 3.7, p87).

### Ethnographic reflections

A significant consequence of my epistemological position (see section 3.2, below) is that I made the decision to include in this thesis a collection of ethnographic reflections on various aspects of my personal experience of OTL. These reflections form the basis of **Chapter 4**, wherein I discuss the particular methodology used in their collation. The



reflections take the form of first-hand reports of being observed and of carrying out observations. The intention behind these inclusions was to provide background and contextual information for the reader, and also to complement the findings of the methods detailed in this chapter.

### **3.2 Ontological and epistemological positions**

My interest in this study was to try to understand and explain how observers reached their OTL decisions in practical classroom situations. A key question concerned the extent to which a researcher can come to ‘know’ how such a process takes place or what it might mean. Does it make sense to say that an event such as a decision can be observed, measured and understood?

Objectivity is much prized by researchers attempting to achieve the status of the natural scientist. However, it might be of limited value when considering human meanings. It may even be impossible: “no human being can ever be completely objective ... we can’t rid ourselves of our experiences” (Bernard, 2011, p328). A qualitative methodological approach based on the subjective experience of observers would, therefore, be consistent with the nature of the subject matter – one which would appear to be shrouded in opacity even for the decision-makers themselves.

I was aware that decisions about methodology and approach to subject arise from the researcher’s ontological view-point and epistemological assumptions. When I considered my ontological position, I concluded that an existential view of reality made most sense to me on a personal level. I believe, with Sartre (1948), that our only given fact was that we exist and everything else proceeded from that, “life has no meaning a priori ... It’s up to [us] to give it a meaning” (p14). Meaning was constructed by

people: “the wholly human origin of all that is human” (Camus, 1975, p24); that social reality is made up of constructions: ‘justice’, ‘love’, ‘effective teaching’...

I do not, of course, choose to live in a state of absurdist revolt. I choose to believe that both the act of research and the endeavour of education itself were worthwhile and of value. I do so because I believe that human action of these kinds – even in the face of a ‘meaningless’ universe – was that for which it was “worth the trouble of living on this earth” (Nietzsche, cited in Camus, 1975, p22). ‘Learning’ helps people to live better lives; ‘teaching’ helps people to learn; ‘effective teaching’ therefore must be a good thing – this was my opinion and belief, and one that I appear to share with a large swathe of humanity throughout history.

From the constructivist paradigm of social research, consistent with this ontological perspective, I accepted that “human beings create their realities in the most fundamental ways” and that “individuals may work together to create a shared reality” (Morgan and Smircich, 1980, p494I). Individuals make sense of their social world through creating and sharing social constructions, and these constructions are “alterable, as are their associated realities” (Guba and Lincoln, 1994, p111). However, I could not accept that there were “no situations other than those which individuals bring into being ... no concrete status of any kind” to social reality (Morgan and Smircich, 1980, p494). It appeared obvious to me that we are acted upon by our social contexts and environments, at least as much as we are free to construct our own realities; just as it seemed equally apparent, following Heidegger’s concept of the “being-in-the-world” (1962), that basic aspects of social reality are emergent from the concrete realities of the physical (and biological) world: “the way the human body [and] the meaningful world of places occur” (Todres and Wheeler, 2001, p5).

For me, then, although the social world is unknowable except through examination of constructed meanings, there was a “‘real’ reality” even if it was “only imperfectly and probabilistically apprehendable” (Guba and Lincoln, 1994, p109). This led me to accept some aspects of the social-realist paradigm of social research. This paradigm holds that the social world is “an evolving process, concrete in nature” (Morgan and Smircich, 1980, p495); which implies that there is a real world and real facts in that world; that there are real skills that can be applied that will have measurable effects and outcomes. As a consequence of accepting this implication, I decided that a degree of pragmatism in the matter of data collection should be allowed to temper my generally interactionist approach.

Given that “the understanding of society is logically different from the understanding of nature” (Winch, 1958, p23), it appeared obvious that specific and appropriate methods of investigation would be required. These methods would seek to avoid what Denzin (1969) referred to as the “fallacy of objectivism” (p926), and instead focus on the lived experiences of practitioners.

I wanted to find out what observers did and why they did it, and the most authentic way I could think of doing that was, primarily, to ask the people themselves – to find out what they said, thought and believed about OTL and what happened during the process. For this reason, it was my belief that the most appropriate methods of primary research would be such qualitative approaches as: undertaking observation; the use of open-ended questionnaires; semi-structured interviews and case study. The element of pragmatism, mentioned above, would manifest itself in the collation of ‘facts’ arising from these methods, and an acceptance that it was possible to make useful statements about a field of study, and that the endeavour could be enhanced via the application of

rigour. For this reason, I have included tabulations of findings drawn from the secondary data, and have commented upon emergent patterns.

As discussed below, I had access to extensive secondary 'quantitative' data arising from the files and records of my research setting. Again, it was my pragmatic opinion that this data could provide valuable comparative opportunities when considered alongside my more qualitative primary findings. Whatever the contingent or technically problematic nature of such data, these facts and figures have meaning within the institution and for the subjects of my interviews and questionnaire; they also arise within the direct context of the particular field of research.

In contrast, perhaps the most qualitative material included in this study was the ethnographic reflections that form the basis of **Chapter 4**. It is my sincere belief that in coming to consider a phenomenon such as OTL, which is for the observer at least partly a deeply subjective experience, it must be of value to take the time to reflect on nature of the lived reality of the phenomenon. The material presented in **Chapter 4** will help the reader to understand the context in which OTL takes place, the personal beliefs and feelings that I bring to the subject as researcher and a fuller appreciation of the context in which my interview and questionnaire subjects offered their responses. My belief was that the consideration of these personal experiences and observations could provide useful data that would complement the findings of the more structured research methods outlined below in this chapter.

I understand the potential criticism that such material might constitute, at best, background information and, at worst, unsubstantiated subjective irrelevancies. However, even here, in compiling my most personal responses to the subject in hand, I have applied a pragmatic approach and attempted to incorporate a degree of structure

and methodological rigour, as will be discussed in the description of my methodological approach to the reflections.

### 3.3 Single-item questionnaires

As a method of data collection, the main advantages of the questionnaire are its “familiarity to users” and, when sent either by post or email, “the fact that it allows them some time to think about their answers” (Muijs, 2012, p146). Questionnaires also have the “advantage of being able to be administered without the presence of the researcher” (Cohen et al., 2007, p317), which saves time and allows data to be collected more widely.

For this research, I used a single-item email questionnaire focussed on observers concerning their OTL judgements:

Respondents	Date of completion
8 Essential Skills observers (not included in Interviews)	04/15

In order to gather the thoughts and opinions of observers from across the Essential Skills team, a single-item open-response questionnaire was sent to them by email. The question was:

*“In your own personal experience, how do you think you actually decide on an observation grade?”*

I presented the respondents with one specific directed question, but allowed an entirely open response. In effect, this made the questionnaire ‘semi-structured’, as I was neither eliciting an undirected general response to the subject of OTL, nor dictating the shape

or content of the response – the design set “the agenda but does not presuppose the nature of the response” (Cohen et al., 2007, p319). In keeping with my epistemological assumptions, this design allowed observers to respond “in a way that they think best ... in their own terms” (ibid); to allow as much freedom as possible to explain the process in their own words and in reference to their own conceptions.

Practitioners were assured that their responses would be anonymous – this, it was hoped, would reduce any inclination on the part of the respondents to include answers they felt ‘should’ be included in terms of the training and management framework of Newbold.

The single question ensured that respondents focused entirely and without distraction on the single issue of their OTL decision-making process and, because questionnaires are always an “intrusion into the life of the respondent” (Cohen et al., 2007, p317), a single question could “minimise the burden” (Bowling, 2005, p342) and increase the likelihood of responses. By adopting this approach, I might have run the risk of receiving very brief responses, as although the single-item technique “can provide valuable information, has the advantage of simplicity, and can be reliable and valid”, this is “at the expense of detail” (ibid, p343). However, in this case, I correctly believed that the subject was one about which these particular respondents would have a lot to say: the average length of response was 215 words.

I chose email as the delivery method as this was the standard form of communication between my colleagues; although, for these purposes, I used my university email in order not to mislead colleagues into thinking the contact was AES work-related. Email had other advantages: “surveys can be done faster than [by] telephone ... the method is also inexpensive” (Schaefer and Dillman, 1998, p3). Van Selm and Jankowski (2006) point out that email questionnaires have also been found to be less of a “burden” for

respondents, an issue for busy professionals. Email entails the “relatively simple task of ... using the ‘reply’ function of their email systems, having inserted their responses into the text of the message returned” (p442). Obviously, this had implications for the confidentiality of the responses. However, as a compromise, I copied the text of the responses over to a file as soon as they were received, where they could be anonymised and the original emails were deleted.

I was aware, from the literature, that an email survey may have poor response rates. Studies have found between 20% (Michaelidou and Dibb, 2006, p293) and 28.5% (Schaefer and Dillman, 1998, p4) for a single contact. My initial response rate was disappointing – two out of twelve contacts - considering the recipients were relatively close colleagues. However, the email had been a bulk mail-shot, possibly resulting in busy recipients over-looking or ignoring an un-personalised approach. Schaefer and Dillman report that “the more attempts made to reach people, the greater the chances of them responding” with response rate rising to “41% for two contacts and 57% for three or more contacts” (ibid). Therefore, I sent reminders to non-respondents; individualising these reminders to specific named colleagues, rather than a second bulk mail-shot. This approach resulted in a final response rate of 70% (eight responses) - which I consider good; exceeding the expected upper range cited in the literature, even for three or more contacts. There did not appear to have been any obvious overriding issues of ‘positionality’ in the uptake of respondents: amongst the 70% who responded are a spread of different ages, gender, race and class, with a very similar profile to the 30% who did not.

My final sample of 8 respondents represented a third of the total number of ES observer-managers at Newbold.

### 3.4 Interviews

Respondents	Date of completion
6 Essential Skills manager-observers (not included in Questionnaires)	14/07/15 to 04/05/16

The interview is a method often chosen by researchers “whose purpose involves understanding more about how individuals think and perceive” (Coleman, 2012, p251); it was, therefore, a method consistent with my epistemological position. I chose to interview because I was interested in understanding “the lived experience of other people and the meaning they make of that experience” (Seidman, 2006, p9). What observers have to say about OTL would provide illuminating insight, and could not be arrived at quantitatively, for example:

“Nobody likes to be judged from the back of the room...”

Here, an experienced observer offers a truth, from her perspective, about the OTL experience. A quote like this can say more about attitudes of practitioners towards OTL than any dry statistic.

I decided to conduct in-depth face-to-face interviews with six observers, out of a total of 24 in the ES department. This was, perhaps, a small sample; however, it was one that was highly relevant to the focus: each subject an experienced practising observer, responsible for creating observation reports in the secondary data. Because of the small sample size, it would not be appropriate to “attempt to generalise” in the same way as “in a large-scale survey” (Coleman, 2012, p251); however, I believed these respondents would provide rich data that might nevertheless prove interesting and valuable. There



were two overarching purposes of these interviews: to collect general data on the lived-experience of OTL, particularly of carrying out observations; and, most importantly, to interrogate the decision-making processes used by observers to reach their judgements.

Due to my personal and professional relationships with my colleagues, I anticipated that my subjects would readily agree to discuss OTL with me. However, I was concerned that I should introduce an element of structure to the discussions – to ensure that the interviews remained focussed and did not degenerate into ‘chatting sessions’. A semi-structured format would allow me “to pursue a specific agenda that had been pre-selected” (James and Busher, 2016, p415) but would not unnecessarily constrain the interviewee in terms of their responses. I did not want to overly pre-determine the direction of the interview, although I had decided on the key questions in advance. In terms of my epistemological approach, semi-structured interviews were “probably the most common type ... [when] working within an interpretive paradigm” (Coleman, 2012, p251), because this approach allows “each participant to respond in their own way... what they want to say becomes as important as what the researcher wants to ask” (Bush, 2012, p79). For this reason, reliability may be more difficult to ensure; however, the method has a claim to authenticity based on its acknowledged ability to “allow for the exploration of the lived experience” of subjects in a particular field and its “great potential to attend to the complexity” of each individual’s story (Galletta, 2013, p9).

In discussion with my supervisors, I devised an interview schedule based around a set of questions linked to and suggested by my key research questions. The semi-structured style of the interview “has its basis in human conversation” (Qu and Dumay, 2011, p246), an approach allowing me “to modify the style, pace and ordering of questions ... to evoke the fullest responses” (ibid); from initial thoughts and factual statements, up to

a consideration of the most important question: ‘how do you arrive at your grade-decision?’

### **Pilot Interview**

Because it was “definitely advisable” (Coleman, 2012, p258) to do so, in order to check that “questions [were] relevant and understandable, and that the interview is manageable within the agreed time” (ibid), I carried out an initial pilot interview with an observer colleague from a different curriculum team. He agreed to take part both as a straightforward interviewee, and as a ‘critical friend’ to comment on the experience of the interview itself.

The pilot interview was recorded with the permission of the subject and transcribed for analysis. I conducted a critical review of my questions and the responses they had received: I considered the effectiveness of the ordering and wording of the questions and the usefulness of the prompts. Tellingly, I had revised the order of the questions during the course of the interview, allowing flexibility according to the ‘feel’ and direction of the conversation. Coleman (2012) suggests that, for semi-structured interviews, there should be a “general consistency in the questions that are asked of each interviewee” but that flexibility is reasonable as long as the interviewer “ensures that the same issues are raised in each interview” (p252).

I noted and considered these changes and, as a result, made alterations to the schedule, omitting one question altogether. The most important change was to reposition the key question “How do you arrive at your grade-decision?” to the end of the interview. My colleague offered only minimal feedback on the interview, but was positive and supportive of several of the decisions I had made: for example, the re-positioning of

the final question: “It really worked being the last question because you have already considered everything and you can pull it all together into some kind of coherence”.

### **Carrying out the interviews**

Following on from the pilot, and the changes put in place, I started carrying out the interviews with the chosen observers (for Interview Schedule, see **Appendix 3**, p220). All were experienced practitioners and observers from one curriculum team within the ES department; all were white-British women over the age of 50. This may seem a highly homogenous sample, but it was generally reflective of Newbold as a whole and, despite superficial similarities, these colleagues were far from uniform in their outlooks, personalities and life experiences. The homogeneity of the sample might have been a concern, reducing the potential “generalizability of the conclusions” (Weiner, 2003, p227), except that I had accepted from the outset that my particular research design was never going to allow strict generalisation. Nevertheless, I decided to include the results of my pilot interview in the sample, as the subject at least differed from the others by being male.

### **Interview Coding**

Coding is the most common method available for storing the results of interviews “in a relevant, usable, and accessible form” (Gorden, 1998, p182) in order to facilitate analysis.

Adapting Glaser and Strauss’ (1967) ‘Constant Comparative Method’, my initial coding was conducted with as little pre-emption as possible, “generating and plausibly suggesting” and allowing “many categories, properties, and hypotheses” (p104) to emerge from within the interviews, rather than directly looking for expected elements. I acknowledged that it would be impossible to entirely disengage myself from past

experience; as Walsham (2006, p321) states “we are all biased by our own background, knowledge and prejudices to see things in certain ways and not others”. However, the freedom of engagement that this method afforded me was useful, and its results illuminating.

Again, in line with Glaser and Strauss (1967), I allowed the content of the interviews to dictate the direction of the coding and this allowed a flexibility of focus; “coding each incident in [my] data into as many categories of analysis as possible” (p106). Later, I was able to “delimit” the field to produce a “reduction in the original list of categories for coding” (ibid, p111). The cross-referencing of codes offered an effective way of comparing and discussing the responses of the interviewees, analysing survey responses gathered elsewhere in my data and bringing these together to construct a thematic account of practitioners’ feelings, experiences, and practices.

According to Gorden (1998, p185), for coding, the question of reliability “asks whether two independent codings of material ... would be the same or whether they would vary grossly”. In order to ensure a reasonable level of ‘reliability’, I undertook to test my coding using an “independent-coder method” (ibid). One of my supervisors produced her own set of codes based on the pilot interview and I undertook a detailed comparative analysis of the two sets of coding. My main interest was in a qualitative comparison; however, I also carried out a numerical analysis of the level of agreement between my coding and that of my supervisor. From a total number of 64, 56 of my supervisor’s codings were in full or close agreement with my own, with superficial differences in emphasis in just 18 of those cases: a percentage agreement of 87.5%. The agreement score is, admittedly, “a crude measure of reliability” (Gorden, 1998, p186) – but might usefully indicate the level of independent agreement between my supervisor’s coding and my own. I considered this a highly positive outcome and

concluded that my initial attempt at assigning code to my interview was along the right lines.

In qualitative terms, I was able to pick up on some very useful omissions and oversights in my own coding. For example, for two specific quotes from the interview, my supervisor coded for 'Fairness' or 'Unfairness'. I had picked up on the same elements but had coded more generally, along with other elements of "Approach to observation". I could see the value of this more specific and emotive label, and recognised that the concept of 'fairness' in terms of OTL, might well emerge as a theme across the interview data as a whole. In this instance, my conclusion was to revise my own coding to include "Fairness".

### **3.5 Analysis of secondary data**

As an insider-researcher, I was able to obtain permission from Newbold's Service Manager to gather secondary data from various institutional files and records, these were:

- **137 observation reports from ESOL, English and Mathematics Curriculum Files (2012-2014)**
- **Annual accreditation and achievement data pertaining to 4336 learners on ESOL, English and Mathematics courses (2013-2014)**
- **Electronic attendance registers for 236 ESOL, English and Mathematics courses (2013-2014)**

The observation reports were of principal interest as the key source for actual OTL judgements and grade-decisions. Accreditation and attendance data were collected for

comparative analysis, being considered, in the literature and within Newbold, key outcome measures of ‘effective teaching’. It was my intention to test common-sense beliefs that ‘effective’ teachers (as assessed by OTL) would have the steadiest attendance and the highest accreditation and achievement rates.

From the observation reports, I was able to draw a large number of examples of real-life judgements: decisions made by experienced observers and the evidence they chose to justify their judgements. Although the data were tabulated and analysed numerically, I believed the real significance of this data arose from its *authenticity*. Comparisons were made for illustration, for what they revealed about how decisions were made and how the judgements were understood by the people who made them. Although I identified patterns and trends in the data, and analysed these where helpful or illuminating, the main purpose was to confirm or contrast with findings within the qualitative data and, therefore, was primarily descriptive.

### **3.5.1 Observation reports**

I had access to extant observation reports for the years 2012/13 and 2013/14 (the completeness of the archive appeared good but cannot be assumed exhaustive). These included follow-up observations required when tutors received Grade 3 or 4 on their first observation. In total, there were 137 observation reports pertaining to 71 observees, as compiled by 24 manager-observers (6 of whom were later interviewed; 8 of whom were the respondents of the observer questionnaires).

The observation reports consisted of:

- A narrative description of the observed teaching session
- A list of ‘Strength’ judgement-criteria (strengths)
- A list of ‘Areas for Improvement’ judgement-criteria (improvements)
- An Observation Grade (see **Table 1.1**, p13)

From each observation report, pertaining to each individual observee, in addition to the subject area, level of tutor and date of observation, the following details were recorded:

- Observation grade (1-4)
- Number of strengths
- Number of improvements
- A verbatim list of strengths
- A verbatim list of improvements

An example of a typical entry for an individual observation report is given below:

Tutor	Level	Date of Obs	Grade	No. Strs	No. Imps	Str - Imps	Strengths	Improvements
Eng14	1	03/12/2012	3	5	7	-2	paperwork	pace
							progress recording	use of ITC
							planning	Explanation of tasks
							Target setting	differentiation
							Learners clear about progress	questioning technique
								learner feedback
								inclusion of all learners

*Table 3.1: Example of a typical entry for an individual observation report*

Tabulating this data allowed me to begin exploring relationships between grade judgements and the criteria cited as evidence for them.

## Numerical representation of observation grades

At the time of research, observation grades were given numerically, from 1 to 4, and were ranked in reported statistics. Below, I have tabulated the results by numerical score, but this requires a note of caution.

It is not the case that there is any mathematical significance to the numerical grade. The value is purely ordinal, “higher values represent more of some characteristic than lower values” (Boslaugh, 2012, p3), however, there could be “no metric ... to quantify how great the distance between categories is” (ibid), and hence no assumption that the periods between grades can be accepted as regular.

I believed it was reasonable to compare grades with, for example, other outcome measures – however, in producing ranges of average grade scores, I offered only a means of grouping results for comparative and illustrative purposes, “not assuming any further properties of the scales” (ibid, p4). I have, therefore, offered no statistical analysis of the comparisons between grades and other outcomes, in terms of correlation or significance, for example. In one particular, however, I may appear to have breached my own injunction in my decision to ‘average’ observee grades over time, for example in **Tables 5.5 & 5.6** (p150 & 152). For each observee in the data set, there were between 2 and 4 separate observation reports (depending on the occurrence of re-observations) for the research period which was also covered by the accreditation and attendance data. That is, that accreditation and attendances rate might relate to a period during which the observee’s OTL grading fluctuated. The presentation of ‘average’ grade scores may be inelegant, but it was an attempt to take this potential fluctuation of OTL grade into consideration – it should not be seen as a suggestion that I believe that the various grades represent a regular mathematical scale.



### 3.5.2 Accreditation and achievement data

Accreditation and achievement rates were routinely recorded, as a crucial part of Newbold's funding formula. Overall statistics were of little use to me, because they were not broken down by tutor or class. I was able to obtain access to the raw data in the form of class lists with 'Learning Outcomes' recorded against the name of each learner. Where learners had been awarded accreditation, this was also listed against their name. Each class was identified by the code used for learning groups on the electronic-register system, allowing cross-referencing to link these learners with their corresponding tutor.

I recorded the results of each class by assigning the numeral 1 to each occurrence of an achievement and/or accreditation. By comparing these results to the total number of learners in the group, it was possible to calculate achievement and accreditation rates for each class. A typical entry for a small class is given as an example below:

Learner	Maths 3	PRO13006	Frqcy	Acc	ILP aim	Acc/rate	Ach/rate
A49246	MATHS ILP	No achievement	1				
E02457	MATHS ILP	No achievement	1				
E03073	FUNC SKILLS	Achieved (aims)	1	1	1		
F00859	FUNC SKILLS	Achieved (aims)	1	1	1		
F01673	MATHS ILP	Achieved (aims)	1		1		
F02586	MATHS ILP	No achievement	1				
F03126	FUNC SKILLS	Achieved (aims)	1	1	1		
			7	3	4	42.85714	57.14286

Table 3.2: Example of a typical accreditation (Acc) and achievement (Ach) data entry for a small class

The rationale for the decision to record a simple numerical tally of accreditation and achievement, rather than any attempt to reflect 'progress', arguably a more valid assessment of teacher ability, was that this would require a calculation based on the

‘distance travelled’ by learners from an initial assessment of their ability at the outset of their course of learning. In practical terms, the achievement data then collated by Newbold did not include initial assessment data. Any attempt to construct a measure of learner ‘progress’ would have required access to tutors’ notes and files, and there could be no way of determining the consistency of these assessments as individual tutors in the different curriculum groups used various different assessment tools and approaches. Therefore, I focused on simple rates of achievement and accreditation, appropriate because, at the time of writing, Newbold was funded purely on the basis of number of qualifications or targets achieved, not on the ‘progress’ displayed.

Having collected accreditation and achievement rates for each class, I was able to calculate average rates for each tutor. I excluded all atypical classes for which special circumstances of context might influence the rates unduly, including: classes for learners with learning difficulties, dyslexia workshops, short courses specifically focussed on individual qualifications, and workplace based courses.

### **3.5.3 Attendance data**

Newbold management has traditionally believed that learner attendance at classes was directly related to ‘effective teaching’. Unlike schools, attendance on adult education courses is largely voluntary. Therefore, learners’ continued attendance might be considered at least an indirect consequence of the qualities of the tutor. In common-sense terms, attendance could be reasonably considered an outcome measure of ‘effective teaching’.

With access to Newbold’s electronic-registers, I was able to calculate an average attendance rate for each observee, and a comparison could be made with their OTL

grades. Only Maths and English classes were considered for attendance rate data because ESOL classes had a much higher and consistent rate of attendance due to the particular cultural and financial imperatives for the learners on these courses. Classes already excluded from the achievement and accreditation data were excluded here as well, and, in addition, the few classes where attendance was mandatory were also excluded: for example, provision linked to the Job Centre.

The electronic-register system presented the following statistics:

- Number of learners starting course
- Number of learners attending overall
- Average attendance
- Potential attendance

The ‘average attendance’ was a simple calculation based on the mean number of how many learners were present at each session throughout the year. ‘Potential attendance’ took account of the fact that not all learners were available for all sessions, for reasons not amounting to actual ‘absence’: for example, not starting until part-way through course, or early withdrawals from the class. Potential attendance was, then, a maximum possible overall attendance on the course. I was able to calculate an illustrative attendance rate for each class by comparing the average and potential attendance values, as follows:

Average Attendance/Potential Attendance = Attendance Rate

### 3.5.4 Judgement-criteria

The strengths and improvements recorded on each observation report constituted judgement-criteria offered by the observer for the grade awarded. These criteria would either be cited as ‘strengths’ or as ‘areas for improvement’. As the criteria constitute the ‘evidence’ for, and detail of, the graded judgements of the observers, it would seem essential to concentrate more closely on the nature of these criteria.

Overall, 55 separate judgement-criteria appear in the observation reports. 27 of these appear on occasions as both strengths and improvements. 19 appear only as strengths and 9 appear only as improvements. Of the 55, 18 appear only once in the data set.

The judgement-criteria collected from the observation reports, were initially in a ‘raw state’, and I needed to create consistent labels for each distinct criterion for two reasons. Firstly, the judgements were written in the observers’ own words with wide variation in the form and phrasing used, even though there were many similarities of focus. Secondly, and partly because of this, when inputting the data from the observation reports, I paraphrased original judgement terminology both for brevity and to apply consistency. This rephrasing was based on my own experiential understanding of the core meanings involved – where these appeared unambiguous. When I was unsure about the meaning of a particular phrase, I recorded it verbatim. Although this short-handing helped to make grouping easier, there was still a considerable variation of wording and potential overlapping of concepts and repetition.

It was, therefore, necessary to bring the various different versions together under uniform labels, in order to facilitate collation. **Table 3.3**, below, demonstrates an example of this process in which appear all the different verbatim wordings collected for judgement-criterion that I chose to eventually label as ‘Activity’.

Verbatim evidence judgement	Initial notation	Final Judgement Criterion
<i>Strengths:</i>		
“Good interactive starter activity	Starter Activity	Activity
“A good variety of learning activities relating to real life”	Real-life activities	Activity
“A wide and very good range of activities to maintain learner interest”	Range of Activities	Activity
“Good range of activities encouraging real learning to take place”	Range of Activities	Activity
“Using small group activities to encourage discussion”	Small Group Activities	Activity
<i>Areas for Improvements:</i>		
“Use activities that are meaningful and demonstrate maths concepts”	Activities for Maths Concepts	Activity
“A good variety of learning activities relating to real life”	Variety of Learning Activities	Activity

“Poor choice of activities”	Activities	Activity
“Starter. This activity did not work.”	Starter	Activity
“Ensure that activities are meaningful and suitably challenging”	Challenging Activities	Activity

*Table 3.3: Examples of conversion from verbatim judgement to final judgement-criterion*

I took care not to smooth out all the differences, allowing a number of potentially overlapping terms to remain, for example: ‘Checking of Learning’ was recorded separately from ‘Assessment’, as this appeared to relate specifically to finding out if learners had understood recent ideas or concepts. Where assessment was identified in more general terms, observers used such phrases as: “excellent formative assessment” or “range of assessment activities”.

Having sorted the judgements under these generic headings, I variously calculated frequency of occurrence: across all observation reports, by grade, by tutor accreditation-rate; by tutor attendance-rate. The intention behind this approach was to interrogate the differences in, and significance of, judgements made by observers according to different contexts.

### **3.6 Ethical considerations**

As part of my accredited course at the University, I attended a session in Advanced Research Methods which focussed on the full range of ethical considerations, especially pertaining to the conduct of research and the protection of participants. I was aware of the fact that, in Sikes' (2004) words, "any research that involves people has the potential to cause (usually unintentional) damage" (p16). For this reason, I have been careful in obtaining permissions; making my role as researcher widely explicit; taking steps to ensure confidentiality; and made efforts to avoid the identification of participants and the institution involved.

During the conduct of my research and the writing up, all my methods and plans were subject to the robust guidance of my supervisors. As required, I submitted a successful application for ethical approval (see **Appendix 6**, p224).

#### **Confidentiality of participants**

All participants were assured that their confidentiality would be protected throughout and, especially, in the final published version. A unique code assigned to each observer mentioned helped ensure anonymity whilst facilitating cross-referencing with accreditation and attendance data.

One caveat must be that, as an exercise in insider-research, it would be obvious to anyone familiar with me that the setting would be the institution in which I was employed. I had to make it clear to potential participants that, whereas I would take all reasonable steps to protect explicit identification, there would be the possibility that our colleagues and former colleagues would perhaps be able to pinpoint 'likely suspects'. None of my participants found this a troubling possibility. However, for this reason, I

decided that no material likely to embarrass or compromise any participant should be included – as absolute anonymity could not be assured under the circumstances.

### **Informed consent**

In the context of insider-research, the question of ‘Informed Consent’ may require special consideration. It has been noted that “even responsible professional educators, [might not] understand what ... they are getting themselves into” (Smith, 1990, p151) when it comes to their participation in research. During insider-research, it is not unusual that a research subject might be a colleague, or even a personal friend. Other considerations, such as sense of loyalty to the researcher, or a social awkwardness, may pressurise the subject to consent, or avoid withdrawing that consent, where a subject unfamiliar with the researcher may feel more free to do so.

Mathison et al, (1993), are much quoted as stating that it is important that “informants may withdraw at any time” (p3); although, in practical terms, it is hard to see how this could be done once the research has been published, for example. Up to a reasonable point, there is nothing in the conditions of the research which would have prevented any participant from withdrawing, or modifying their consent, if they felt that anything had given them cause for concern; and I made it clear to all that this was the case. At the time of writing, no participant has so decided.

There is also the question of the power status of the relationships involved. Bogdan and Biklen (2007) suggest that ethical issues may arise when studying colleagues, peers or people over whom you have some authority in a setting” (p57). I did not interview any colleague in a role subordinate to my own; however, in the case of the inexperienced observer mentioned in **Chapter 4** (section 4.3.2, p113), I combined an informal mentoring role alongside that as researcher. I assumed that this arrangement



represented an agreeable quid pro quo, but upon reflection, I could see that my colleague may well have felt unable to refuse. Taylor (2011) suggests that “the most constant form of effective ethical management” has been to offer participants “the opportunity to review their transcripts, allowing them to add or to revoke anything that has been said in the interview context, and to view my written work in which they are cited and interpreted prior to submission for publication” (p16). In the light of this suggestion, I shared with the newly-appointed manager a draft of the material I proposed to include pertaining to her participation. I was open to making changes or, if she decided to withdraw consent entirely, to omit it. Her response was to reconfirm her consent, with the exception of one adjective which she thought was unnecessary. The removal of this word did not change the account in any substantive way, and I was happy to comply.

### **Access to secondary data**

In order to collect data, I required access permission from Newbold’s Service Director. At a face-to-face meeting, I explained the purpose of my research and the specific use I intended to put the data in the records to; shared a copy of my ethical approval form; and explained the procedures I would put into place to ensure the confidentiality of all material relating to staff members and learners. As a serving manager at Newbold, I already had routine access to much of this material, so it was a matter of extending permission of use to include my research purpose rather than granting new access. The Service Director granted permission in general terms, but with the proviso that I obtained agreement from the three ES management teams, specifically on the question of access to tutors’ observation reports.

Two of the ES management teams gave their consent without further discussion. The third, however, having debated in closed session, were initially minded to decline access

to their archives. The main concern was sensitivity around an 'outsider' accessing tutors' reports, when some of them had received low grades, which speaks of the considerable level of sensitivity surrounding the entire subject of graded OTL.

I asked for and received permission to meet with the management team to directly address their reservations. Again, I explained the purpose of my research and interest in the observation reports; distributed copies of my ethical approval form and emphasised the way in which I would ensure anonymity. I stated that I would be happy to address the concerns of tutors in an email and would not access the records of any tutor who wanted their records excluded. This satisfied the management team and they were happy for my research to proceed.

I had not anticipated that any staff member might be concerned about my accessing their records, especially as I already have such access in my work role. However, I was grateful to the team for raising this issue, as it gave me the opportunity to address any such worries in advance rather than leaving myself open to complaint at a later stage. I distributed an email to all ES tutors explaining my plans and asking for any questions or objections to be addressed to me, either directly or through line management, attaching my ethical approval form. As of the time of writing, I have received no objection.

### 3.7 Summary of methods

#### Method details

Method	Subjects	Dates
Questionnaires	8 Essential Skills manager-observers	04/15
Interviews	6 Essential Skills manager-observers	14/07/15 to 04/05/16
Analysis of secondary data	71 Essential Skills tutors:  [137 observation reports (2012- 2014)  236 classes (attendance data) (2013- 2014)  4336 learners (achievement data)  (2013-2014)]	09/14 to  03/15

*Table 3.4: Method details*

## How were the research questions addressed by the methods?

Method	Research Questions
Observer Questionnaire	<p><b>RQ2: How do observers explain their OTL decision-making?</b></p> <p>Question specifically focussed on explaining decision-making.</p> <p><b>RQ3: What elements of ‘effective teaching’ do observers focus on in evidencing their OTL decisions?</b></p> <p>Expected that explanations may include details of elements considered important by observers.</p> <p><b>RQ4: How do observers’ judgements relate to other potential outcome measures of ‘effective teaching’?</b></p> <p>Possible explanations might include comparisons with other methods of measuring ‘effective teaching’.</p>
Observer Interviews	<p><b>RQ1: What do observers believe about the OTL process?</b></p> <p>Interview will directly address observers’ beliefs..</p> <p><b>RQ2: How do observers explain their OTL decision-making?</b></p> <p>Interview will ask observers to try to explain their decision-making.</p>

	<p><b>RQ3: What elements of ‘effective teaching’ do observers focus on in evidencing their OTL decisions?</b></p> <p>Interview will ask for key elements; respondents may mention others in reply to questions.</p> <p><b>RQ4: How do observers’ judgements relate to other potential outcome measures of ‘effective teaching’?</b></p> <p>Respondents may refer to other outcomes and make comparisons.</p>
Analysis of secondary data	<p><b>RQ1: What do observers believe about the OTL process?</b></p> <p><b>RQ2: How do observers explain their OTL decision-making?</b></p> <p>Analysis of data from observation reports will reveal what observers actually do as opposed to what they say they believe about OTL and how they explain their decision-making.</p> <p><b>RQ3: What elements of ‘effective teaching’ do observers focus on in evidencing their OTL decisions?</b></p> <p>Analysis of data from observation reports will demonstrate which elements of ‘effective teaching’ observers actually cite in their decisions. The criteria can be compiled and compared across grade levels.</p>

	<p><b>RQ4: How do observers' judgements relate to other potential outcome measures of 'effective teaching'?</b></p> <p>Achievement and accreditation data and rates of attendance can be compared to grade outcomes in order to test the relationship between these factors.</p>
--	--

*Table 3.5: Methods used by research question*

## **Chapter 4 – Ethnographic Reflections**

### **4.1 Introduction**

OTL is a human process involving the judgement of one person about the professional abilities of another. For anyone who has not personally experienced being on either side of this process, it might be difficult to fully appreciate the concerns, practicalities and the anxieties involved. I must also acknowledge that, as a practitioner of more than twenty years' experience, I cannot pretend that I do not have my own deeply felt personal feelings about the subject. This chapter, therefore, represents my attempt to make the experience of OTL relatable for the reader, as well as to bring my own feelings and professional experience to bear on this research. To this end, I intend to share some reflections on personal experiences of aspects of OTL.

I consider these reflections to represent exercises in ethnography, as opposed to unstructured reminiscence or anecdote. The term 'ethnographic' suggests a "strong emphasis on explaining phenomena within their social settings" (Anderson with Arsenault, 1998, p128), in this case Newbold, and involves a form of analysis which "emphasises description and explanation" (ibid, p129) rather than the quantitative analysis of traditional positivist approaches. Consistent with my epistemological approach, I have introduced methodologies that have helped me to "interrogate my own experiences" (Sikes, 2008, p 154). An ethnographic methodology is one which will admit "into the research frame, the subjective experiences of both participants and investigator", perhaps providing a "depth of understanding lacking in other approaches to investigation" (LeCompte and Goetz, 1982, p32).

The two methods I adopted were: auto-ethnography, in order to reflect on personal experiences of being observed and of carrying out an observation; and observing

participation, in co-observing in classrooms with both an experienced and a newly-appointed observer-manager.

In the sections below, I have collated my reflections on the following experiences:

- Personal experiences of being observed: including an in-depth analysis of my approach to preparing for observation; and my interpretation of the feedback received.
- Personal experience of carrying out an observation
- Observing participation of a classroom observation by an experienced observer-manager (co-observing)
- Observing participation of a classroom observation by an newly-appointed observer-manager (co-observing)

In order to avoid the reflections becoming merely anecdotal, I was keen to ensure that I was, as far as possible, “systematically” analysing my “personal experience” (Ellis et al., 2011). In each section below, as well as reporting on the results of my reflections, I outline the methods I put in place in order to ensure some degree of rigour.

## **4.2 Auto-ethnography**

An auto-ethnographic approach “seeks to describe and systematically analyse personal experience in order to understand cultural experience” rather than “hiding from these matters or assuming they don't exist” (Ellis et al., 2011, p1). Denzin (1969) pointed to the positive value of such an approach: areas of study such as OTL involve capturing the “unfolding meaning” of “behaviour of both the covert and overt variety” and



analysis establishing “what is taken for granted and what is problematic”. This can be effectively achieved by “making the self a central object of study” (p925) when that “self” is a practitioner involved in the setting and practice of the subject matter.

Specifically, I planned to focus on my own experiences of two aspects of OTL: being observed and observing another practitioner.

#### 4.2.1 Experiences of being observed

In this section, I will reflect on the emotional, behavioural and professional impact of the personal experience of two annual observations, as detailed below:

Observation 1	Observation 2
Date: 09/05/14	Date: 24/02/15
Lesson Type: GCSE session for adult learners	Lesson Type: GCSE session for adult learners
Observers: Line Manager; Co-Observer (experienced Observer)	Observers: Line Manager
Grade outcome: Grade 1 (outstanding)	Grade outcome: Grade 2 (good)

*Table 4.1: Details of Observations 1 & 2*

Whilst acknowledging that any degree of true objectivity would be impossible in an exercise of this kind, in order to undertake “observation in some methodical way”, I wanted to attempt to achieve an element of what psychologists refer to as “structured

introspection” (Carrette, 2007, p58). In order to create a distancing effect between the different roles of practitioner and researcher, and to bring focus to the process of reflection, I devised a structured self-questionnaire consisting of a set of questions designed to capture my comments and insights (see **Appendix 2**, p218, for list of questions), although I understood that the choice of these questions would inevitably reflect the areas of concern and assumptions already formed from my previous experience.

In completing the questionnaire, in order to retain a sense of immediacy, I tried to keep my responses brief and spontaneous, with as little ‘pre-editing’ as possible. In this way, I hoped to capture my thoughts more ‘authentically’, as it were, before the researcher side of me could organise the data in line with expectation. I was aware that this ‘authenticity’ might be illusory, however this approach helped me to focus on my experiences in a way that that I believed was useful – and offered some separation between the setting of the questions and the collection of responses.

### **Emotional Impact**

From the first, my emphasis is on the emotional impact of the experience. "A sinking panic" is reported, "a sick feeling". Whilst acknowledging the routine nature of the OTL, it "doesn't feel routine" to me. From the moment that the observation date is set, there begins an "anxious cycle of doubt and denial".

As Observation 1 approaches, I report "more panic"; I am suffering from "performance anxiety". At this stage, I have expressed no resentment towards either the process or the observers, rather there is self-recrimination: "all the negative feelings focused inwards".

Looking back on the observation, there is a sense of relief, "I can relax for another year". OTL is acknowledged as something to be "survived".

For Observation 2, the following year, I decided to cut down on the amount of additional preparation undertaken for the observed session. In effect, I would make a stand against the “game playing” previously engaged in, and do no more than the routine preparation expected of a “normal” weekly lesson.

However, in practice, it was not as easy as might have been expected. It felt like “something was missing” and I found myself lying awake at night thinking of potentially impressive activities to fit into the session. I knew this agitated obsession with the session was not the normal run of things. It was the consequence of “observational anxiety” (Brown et al, 1993, p23).

The lack of extra preparation meant that my anxiety had actually increased. As the session approached I felt a great deal of foreboding. I felt like I was “flying without a parachute”; I was “naked and exposed”. I could not sleep the night before the session and I was consequently tired by the time it came around.

The actual result of the observation came, initially, as a relief, but soon was seen with a certain shame – I felt as if I had “let myself down”. I was left feeling that the next year, I would not repeat the experience, and would be back to my safety net of “putting on a show” in order to “return to form”.

### **Behavioural Impact**

For Observation 1, my normal work practice was altered in direct response to the impending event. Greater emphasis was placed on preparation for this session, taking up much more time than usual. Throughout, the emphasis was on presenting a positive impression to the observers.

Knowing that the ‘invisible facilitator’ teaching style was privileged in the OTL process, the lesson plan exaggerated this aspect – an example of “playing the game”, as researchers have found elsewhere (Woods and Jeffrey, 2002, p102).

I took deliberate strategic advantage of the observer's offer for me to choose the class to be observed. Activities were planned in order to "demonstrate the abilities that I know the observers are looking for". This is in line with O’Leary’s (2013) findings of an “openly cynical” and “pragmatic response to the use of graded OTL” (p709). The observation was approached as if it were a theatrical performance: I was “sick with stage fright”. Woods and Jeffrey (2002) also found that teachers approached their observations in this way, with the most highly graded being those who "know how to put on a performance" (p102).

The following year, for Observation 2, I made the decision not to over-prepare for the observed session, and this paid dividends in terms of minimising the effect of the pre-observation period on my normal routines and focus on learner needs.

However, a few days before the observation, I felt that I may “have made a mistake” and that it might be necessary to prepare a starter activity that would be novel and impressive: basically a “gimmick”. This consisted of arranging for the learning aims for the lesson to be “shared” (an observation essential), via text message to each learner simultaneously – this was to tie in with and illustrate lesson content about multi-modal communication. This took a lot of time to set-up – and was almost entirely motivated by the need to “show off” to the observer.

As a result, the session was not as over-prepared as before – but still considerably more than for a routine teaching session.

**Professional considerations:**

As well as personal considerations, I was also concerned about the impact of the OTL process on the teaching and learning experience itself. My responses revealed that my pre-occupation with the observation had potentially detrimental effects for the quality of delivery:

- Lack of learner-centeredness: Focus on the learners' needs was relegated for the duration of the OTL process. Instead, the priority was "to show myself off to the best of my ability" whilst "matching the criteria I know they are looking for".
- Diversion of time and resources: Greater time and effort was lavished on the observed lesson, at the expense of other sessions and duties, for example, extra time was spent ensuring paperwork "is present / up to date". The entire OTL process was termed "a time consuming farce". In Observation 2, the last-minute inclusion of specially devised material and activities diverted several hours from other tasks.
- Efficacy of the OTL process: to the question, "How will the feedback affect your future teaching?" The answer was simply: "It won't". Overall, "I didn't think I learned anything", the feedback was not constructive, mostly being praise. Ironically, some of the praise given was "specifically picking up on efforts made only to impress the observers". Overall, there was a perceived lack of "benefit to my learners"; a feeling that "the whole thing ... diverts me massively from my work for at least two weeks".

## **Observer Feedback**

### **Observation 1**

My Line Manager made it clear that there had been a disagreement between the two observers as to the grade-decision. The Co-observer had suggested a Grade 2 (good), whilst my Line Manager had opted for Grade 1 (outstanding). She had “argued the case”, and they had eventually agreed on the higher grade.

The disagreement centred on the question of teaching technique: the Co-observer said that he had “not seen [me] doing enough actual teaching to judge”. My decision to present myself in the facilitating role, that was currently in paradigmatic favour, had led him to feel that I had not demonstrated teaching ability. My Line Manager had to challenge this view by explaining the prominence of the ‘facilitator’ approach.

This negotiation at the heart of the grade-decision brings into focus the constructed and context-dependant nature of the decision-making process.

I was ambivalent towards this decision. Personally satisfied that my tactic had resulted in a high grade, I was also keenly aware of the irony that the Co-observer had wanted to mark me down due to my abandonment of my usual preferred style.

### **Observation 2**

In this observation, I was awarded a Grade 2 (good). In feeding back, my Line Manager appeared to be as disappointed as myself, if not more so. She was at pains to excuse my lower than expected result: doing so in reference to my previous record and in accordance with her understanding of my “real” abilities: “I know you are a grade 1 teacher...” The nature of the class and venue, the timing of the session, the technological failure of my otherwise “brilliant” opening gimmick – all are called upon

to excuse my inability to achieve my “rightful” grade. I did not tell my Line Manager that I had not prepared for the session in the usual over-extended way.

I was left wondering if my disappointing, but more than adequate, grading more closely represented my actual level of effectiveness under normal circumstances. Perhaps the grading structure is undermined by an over-inflated idea of what a tutor should be capable of. The two week notice period might raise unfair expectations of what can be achieved in the classroom; an expectation of ‘something a bit special’, unsustainable in day to day practice.

I was concerned that, given this inflationary tendency, a routine session, visited without notice, might actually fall into the category of ‘unsatisfactory’. Normally, I have only a few hours to prepare and reflect on my teaching sessions and perhaps that is the kind of teaching that should be observed. It would be necessary to achieve a recalibration of the standards expected of observed practitioners, but might result in a clearer picture of day-to-day teaching, as well as a reduction of stress and the effects of over-preparation.

### **Reflections on observer’s judgement-criteria**

In this section, I shall share my reflections on some of the judgement-criteria presented as evidence for the observer’s grade-decisions in the two examples discussed above, which, I believe, were influenced by decisions I made strategically in preparation for the observation. I will attempt to interpret how each example may have been selected by the observer.

### Observation 1 – Grade 1 “Outstanding”

Observer’s Judgement	Comment
<b>Independent Learning</b>  “Excellent focus on asking learners to think for themselves”	This is a teaching method I often adopt, but here I took this to an extreme. A perfectly valid technique but one that I could not use to the extent displayed here on a routine basis because I would not have time in the course to get through the required curriculum content.
<b>Classroom Management</b>  “Learners worked in pairs and in small groups to share their own assessment of what the exams will require”	A deliberate decision was made to move learners around the classroom and mix them up, based on my understanding of the current paradigm. I focussed the content of the session on sharing learner knowledge specifically to demonstrate this to the observers.
<b>Questioning Technique</b>	Less conscious and perhaps more indicative of habitual practice. However, I had planned the inclusion of a “gimmick”: asking learners to provide questions to fit a series of answers I gave them. This conscious inclusion may have influenced the judgement.
<b>Paperwork</b>  “Excellent course file”	My weak-point. I spent considerable time preparing my course file and making sure it was complete. None of my other class files were anywhere near this standard.

Table 4.2: Reflections on observer’s judgement-criteria (Observation 1)



## Observation 2 – Grade 2 “Good”

Observer’s judgement-criteria	Comment
<b>Subject Knowledge</b>	I made no special effort or conscious decision to display this. The previous year, this was not mentioned – perhaps because there were more strengths to mention?
<b>Activities</b>  “Innovative – lesson objectives by text – text analysis”	The “gimmick”. I received considerable credit despite the activity falling flat (the texts came through late and spread out rather than on cue). This was, as discussed, a result of ‘observation anxiety’ and was a late inclusion.
<b>Pace</b>  “A little slow at times”	The sole improvement. A direct consequence of the relaxation of planning time put in to the lesson preparation. Perhaps this session more fairly represented the usual flow of my teaching?

*Table 4.3: Reflections on observer’s judgement-criteria (Observation 2)*

### Thoughts:

Arguably, my ability to ‘finesse’ high grades indicates a higher than average understanding of the requirements of the process. My skill in being able to produce and display the elements of ‘effective teaching’ was based both on my teaching experience and on my knowledge of the criteria. Does this represent a deception or indicate my ability as a teacher?

Based on my meditations here, the question remains as to the efficacy of OTL as a measurement of ‘effective teaching’. Given that explicit presentational ‘gaming’ can factor in the outcomes of some observations, are those practitioners who fail to achieve high grades just more honest in their unadorned presentation during observations, less

adept than others at making their skills and knowledge explicit, or are they genuinely less able teachers?

Without access to the internal processes of other practitioners, it would appear impossible to assess how common the approaches I discuss here are throughout the profession. Given the high-stakes nature of OTL, it is debatable whether practitioners would acknowledge their use of such strategies, even if they were asked. However, I have hopefully demonstrated that the “game playing” noted by other researchers, and described above, may be a factor affecting grade-decisions.

#### **4.2.2 Experiences of carrying out observations**

My first experience of observing and grading another practitioner came in 2008, shortly after assuming managerial responsibility. I felt unprepared for the duty, apart from a sense of belief in my own experience as a teacher – that I would ‘know good teaching when I saw it’. I was supported in my first observation by an experienced co-observer, who acted as mentor. We did not discuss any criteria or grading prior to the observation, but I had read a summary of the then current Ofsted Common Inspection Framework. I did not find this particularly helpful in terms of specifics, but it provided me with general headings and a structure within which to organise my thoughts. Following an hour watching the lesson in question, we withdrew to a private room to compare notes.

My co-observer asked me to state the grade I had in mind. This, I felt, put me on the spot. However, in the spirit of the exercise of mentoring, I understood that my first opinion would be used to indicate the extent to which my judgement could be ‘trusted’. Happily, although not especially instructively, we agreed on the grade – and this was

seen as ‘evidence’ that I was on the right track. I remember feeling relieved that I had survived the experience with my self-esteem intact! However, there was another core concern, I was anxious that my grade should reflect the quality of teaching I had observed in that classroom. I wanted to be ‘fair’ as well as ‘right’, in terms of agreement with the experienced judgement.

In order to present an account of what it feels like, personally, to carry out an observation, I share below a written record of one such OTL session. Taking advantage of my insider status within Newbold, I used a regular annual observation of a tutor as an opportunity to capture my experiences of observing:

<b>Observation 3</b>
Date: 24/11/14
Lesson Type: Essential Skills Workshop for adult learners
observer: Tutor’s Line Manager (researcher)
Grade outcome: Grade 2 (good)

*Table 4.4: Details of observation 3*

I did not inform the tutor, beforehand, that I was conducting research during her observation – so as not to further increase her stress level; and because the focus of this exercise was on myself as observer, rather than on her performance as teacher. I informed the tutor after the observation that I would be recording my thoughts and feelings as an observer for my research, but would not include any details specific to her. She was happy to consent to this.

I came to the conclusion it would be too difficult to juggle a systematic schedule of questions to think about during the observation, as I was there primarily to carry out the OTL. So, the method I adopted was to record my thoughts and feelings on paper as immediately as possible during and, to a lesser extent, after the experience of the observation. I would then reflect on these notes in my analysis. It was hoped that the ‘immediacy’ of my responses would provide sufficient distance between my twin roles as observer and researcher. I would, in effect, be presented with my notes as a documentary ‘resource’ to analyse and comment upon, rather than just recording my anecdotal account.

#### **An account of a classroom observation (based on verbatim notes)**

Situating myself to the side of the class, I made myself a little ‘observation station’ out of my files and paraphernalia – a base from which to carry out my observation. I felt comfortable there, but a stranger in the classroom. I had never taught in this room nor met with these learners before.

My main focus was on the tutor: the way she handled the learners, spoke to them, introduced the activities. I was aware of getting an immediate ‘impression’ of the tutor, which was reinforced throughout the session. I had to look at the paperwork, planning and so on, and I could do that after a while, because I became confident in my ‘first impression’, and was fairly sure I was not going to miss any telling detail.

A good deal of the grade-decision was based on this initial intuitive ‘feel’. It just ‘felt’ like a well-managed, professionally put together session was taking place. Secondary evidence that learning was taking place was found in the learners’ work in their files; the way they responded to their tutor; frequent mentions of ‘marking’ and ‘progress’. The learners seemed confident and comfortable with the tutor.

A 'nugget' of detail emerged when the tutor broke up a pair of learners who were talking 'off-task'. I felt relieved, I had got a detail! This would give me an anecdote to include in my report: evidence of effective classroom management. I was actively looking for clues and ticking boxes. Further boxes were ticked by talking to the learners: were they happy with their class? Did they feel their needs were being met? And so on.

My grade judgement was mainly based on the holistic feel I got almost from the start of the class and then justified by reference to the pieces of detail pulled together like detective work throughout. Strengths and improvements were not explicitly collected in the session – these would be decided during the writing up of the report.

### **Thoughts:**

Reflecting afterwards on these notes and this experience, I was struck by how automatic the grade-decision seemed – appearing in my mind very quickly and being maintained consistently throughout the observation. My gathering of supporting evidence and detail was mechanistic in nature; secondary to the decision which I had more or less already made. The enthusiasm with which I leapt upon the telling anecdote of the tutor separating the learners is indicative of at least one priority: I had a report to write and was anxious to find detail with which to fill it.

It seemed clear that my focus during the observation was on tutor performance and that the key criteria of 'learning taking place' was itself largely unobservable – perceivable only through secondary 'clues'. That strengths and improvements were not determined, consciously in any case, during the session clearly indicated their secondary status when compared to the holistic grade-decision.

### 4.3 Observing Participation

According to Bailey (1994), observation is valuable when it is necessary “to study in detail the behaviour that occurs in some setting or institution” in order to construct a “comprehensive in-depth picture” of that “behaviour in [that] particular setting” (p242). Observation offers the opportunity to “gather ‘live’ data from a naturally occurring situation” (Cohen et al., 2007, p396) as opposed to relying on second-hand accounts.

In accordance with my epistemological position, I decided to utilize the ethnographic method of ‘observing participation’; an approach in which the researcher, fully involved in the world of the research itself, “becomes the data collection instrument” (Anderson with Arsenault, 1998, p141). Observing participation is a development from the better known technique of participant observation, in which the observer has entered the field of research from the outside. Marek Kaminski (2004) explains the limitations arising from the element of distance and objectivity traditionally conferred by the participant observer’s outsider status. They are, he suggests, “not as affected emotionally by the events” (p7) which means they might observe without really understanding the deeper significance of a situation – they lack the “experiences that can stimulate one’s understanding of insiders’ problems” (ibid). In contrast, the observing participant, as an insider-researcher: “undertakes field research as if he or she was a researcher” (ibid); they really are participant, with full experience and understanding of the situation, but simultaneously take on the role of researcher in order to systemise their analysis.

In this mode, I participated in two routine classroom observations, paired with colleague observers, in order to explore their decision-making processes, and to compare these with my own understanding of the same experiences. I took advantage of the advent of a newly promoted manager within Newbold to allow a point of

contrast between the two accounts offered: one highly experienced observer, and a manager carrying out her first observation, as detailed below.

<b>Observation 4</b>	<b>Observation 5</b>
Date: 13/03/15	Date: 10/02/16
Lesson Type: Essential Skills Workshop for adult learners	Lesson Type: Essential Skills Workshop for adult learners
Observer/s: Tutor's Line Manager (experienced observer); Co-observer (researcher)	Observer/s: Tutor's Line Manager (first- time observer); Co-observer (researcher)
Grade outcome: Grade 2 (good)	Grade outcome: Grade 1 (outstanding)

*Table 4.5: Details of observations 4 and 5*

I attended each observation as an observing participant, playing the familiar role of the co-observer. In each case, the tutor being observed was aware of, and happy with, my status as researcher. As during any routine paired-observation, it was the main observer's responsibility to produce the written observation report. My role was to share my own thoughts on the observed session, discuss and agree the grade-decision and provide my own notes to assist with the writing of the observation report. I made copious notes of the observed session and arrived at a grade-decision. Following the period of observation, I listened to and make notes of the main observer's reasoning and judgements.

#### 4.3.1 Co-observing with an experienced observer

This observation, of an experienced Essential Skills tutor, took place during a mixed-ability workshop session, with 8 learners - all second-language speakers, all women, culturally diverse. Teaching and learning focused on skills needed for Level 1 and Level 2 accreditation. For various reasons, my opinion was that the session merited a Grade 3 (requires improvement). Despite my reservations, the observer decided to award this tutor a Grade 2 (good). She was aware that this judgement was not altogether justified in terms of what had been demonstrated in the session:

“I think it is a Grade 2 but not a zingy 2 - I don’t want to give a 3 because it seemed to suit the learning style of the group.”

The observer’s previous knowledge of this tutor was clearly a factor in the grade-decision:

“I don’t want to give her a 3 because I know what she is capable of.”

I gave my opinion, that I would have given the session a Grade 3. The observer admitted that it was not a good session; however, she would not be happy with anything less than a 2:

“One of the worst things to be accused of is not caring. I know [the tutor] is not a 3. We should be empowering, building up people not discouraging.”

Therefore, the grade was set at 2, although we agreed that it would be verbally presented to the tutor as a “low 2” – a ‘*de facto* 3’ - requiring much improvement.

The observer revealed anxiety about the potential for bias in her decision-making: “I do wonder if I was allowing my prior knowledge of [the tutor] to colour my



judgement". She acknowledged the wide range of conflicting elements that go towards arriving at a grade, and rationalised her decision to err on the generous side by suggesting that the criticisms omitted from the observation report could form the basis of an action plan: "I will use some of the more detailed and specific stuff, in my action plan." This compromise allowed her to avoid committing the criticisms to the official report.

What was left unsaid was that an action plan had little currency when a tutor has already been awarded a Grade 2. Because, although in principal even highly-graded observees might have some improvements to work through, in practice action plans were rarely followed through unless the observee had 'failed' their observation and required re-observation.

### **Thoughts:**

Analysing this experience provided insight into the highly subjective and consciously constructed nature of OTL. It was clear that a grade-decision could be highly context-specific: what mattered most to the observer was her personal knowledge of, and high regard for, the ability of the tutor. This acted as a corrective lens through which she adjusted her focus on the particulars of what was a disappointing teaching session. The observer was determined to protect the tutor from a disappointing (and potentially humiliating) grade judgement.

Institutionally within Newbold, observation reports were considered objective QA evidence; they also provided official data to be reported to Ofsted, for example.

However, it is clear that, in this instance, the observer's approach to creating the report was to provide a narrative which deliberately softened the impact of some criticisms,

omitted others and placed undue emphasis on positive aspects – especially those of paperwork and planning in the tutor’s records.

The observer here was sincere in her belief that this tutor was generally a “good 2”.

There was a significant gap between that judgement - based on previous knowledge of the tutor over time, of knowing what she was capable of – and the specific outcomes of a ‘snap-shot’ observation. The question raised was whether observers should make grade-decisions based solely on what they see, or should what they know also be taken into consideration?

#### **4.3.2 Co-observing with an inexperienced observer**

This second experience of observing participation involves co-observing with a newly-appointed ES manager. As a first-time observer, she was keen to have the support of a co-observer and was also happy to accommodate my researcher role.

During the pre-observation conversation, the observer demonstrated an awareness of the fact that OTL serves a variety of different purposes. Chief amongst these was “ensuring standards across the board”: a concern both with the quality of teaching by individual practitioners and a moderated standard across the whole provision. This suggested a belief in achieving standardised results: “we have to make sure all the standards are the same”. She believed that OTL might be an effective method by which to “prove” and “evidence” these standards. She recognised, however, a tension between the QA purpose and the CPD role which she believed should be prioritised. She was apparently torn between believing that observations “are good in that they help to ensure standards” and a critical awareness of the negative impacts of grading.

As well as intending to draw upon her experience as a practitioner and as the subject of many observations, deriving key observation criteria from those experiences, she also believed that there were external “standards” that existed and that were applied, although she was unsure as to their origin: “I’ve always assumed they come from Ofsted; they’re the criteria Ofsted uses”.

She admitted to finding the process of grading mysterious: “I’ve never been aware of what decides a two or a one or a three... I’ve never understood where they’ve got it from”. She expressed a contradiction in trying to explain that judgements come from both “the criteria that Ofsted uses” and that “it also has to be subjective – based on your knowledge of teaching”.

She was both confident in her experience as a classroom practitioner, and concerned about her ability to justify experience-based decisions in light of some presumed hidden standard: “I can look at it and make a decision but whether I can evidence it properly... to the standard that’s required...” She was prepared to trust her “feelings” but worried that she would go “really off-kilter”. She was torn between a trust in her experiential understanding and the worry that there was an objective, or at least normative, standard that governs the application of criteria to the decision-making process, and of which she might have been unaware.

Following the conduct of the observation, the observer’s initial judgment was that it “was a good lesson”. Immediately, she then attempted to rationalise that judgement in terms of criteria: it was “well structured”; “linked to previous learning”; “good pace”.

She acknowledged that personal knowledge of the tutor and her learners would be an element in her decision-making. Commenting on the tutor’s handling of a challenging learner, she revealed that the tutor “does beat herself up unnecessarily about it”.

The observer also demonstrated subjectivity in choosing which judgement-criteria to consider as important or to ignore. She clearly did not credit the use of the Virtual Learning Environment (VLE) very highly: “I mean, is it essential? Do you have to mention it?” According to Newbold’s Annual Review, the use of the VLE was a priority area at that time. However, she was reluctant to see the tutor’s lack of uptake in this area as an area for improvement.

My own opinion of the observed session was that it was excellent; I would have had no difficulty in awarding it a Grade 1 (outstanding). When asked about her own decision, however, the observer became quite anxious. This decision was clearly much more difficult than coming up with a list of judgement-criteria, which appeared to give very little concern. The connection between the judgement-criteria and the grade-decision was clearly not as direct as might have been anticipated.

This difficulty appeared to contradict her previous confidence. She mentioned three major barriers between her instinct as a practitioner and her ability to arrive at a grade, the first of which was a point of procedure: the belief that an observation will only qualify as Grade 1 if it is entirely free of areas for improvement. This was a widely held belief amongst practitioners in the service; in fact, almost a folk myth, as an appeal to verbal history acknowledges: “it’s always been said that if you have one or two improvements, you can’t go for a 1”. This assumed block allowed her to assert that the grade for the lesson must be a 2.

However, she was clearly not satisfied with this decision. After a long pause, she revealed the second barrier to her decision-making: the fact that as an observee, she had struggled sometimes to understand why she had received a particular grade:

“I’m going to struggle here because... I’ve had a 1 – and I have absolutely no idea why I got a 1. I had the feedback, but I still don’t see what was a 1 about that. What is the difference between a 2 and a 1? This is where I’m going to need help...”

This quote, considerations of modesty notwithstanding, clearly illustrated the mystery surrounding grade-decisions.

The third barrier alluded to in this discussion revealed itself only after lengthy inner struggle. When asked to respond purely on instinct, putting the task of report-writing to one side, she was unequivocal: the class was “excellent” and the grade should be a 1. Her concern was how she would be perceived by her new management colleagues:

“What are they going to say? ‘Her first ever observation and it’s a 1!’ ... I’m worried that giving a 1 in my first observation; they’re going to turn around and say: ‘she doesn’t know what she’s talking about!’”

Having previously acknowledged the high-stakes nature of the OTL decision for the tutor, here, she revealed the high-stakes nature of the decision for herself as a new manager. She was concerned with the external element of peer-acceptance to the point to which it might influence her grade-decision. It was only the fact that I had also decided on Grade 1 that allowed her follow her instinct.

**Thoughts:**

This new observer's journey through her first observation was characterised by doubt: about the procedure of decision-making itself; about her ability to make a "correct" judgement. She did not doubt her experience as a teacher, or her knowledge of teaching craft – only the calibration of the grade-decision worried her, along with the legitimacy of her 'inexperienced' opinion. She, therefore, demonstrated the importance of her own teaching experience and knowledge of service practices and traditions – her 'Wisdom of Practice' - as the basis of her judgements. She explicitly cited her experience, both as a classroom practitioner and as a subject of many observations, as a major factor in her decision-making. She also brought external factors into the mix: prior knowledge of the tutor and learners; rumours and 'folk-wisdom' surrounding the grading of observations.

The observer's preconceptions revealed just how much was hidden from the practitioner who is the subject of it. She was unsure where the standards came from; she did not know what observers talked to the learners about because tutors were always obliged "to be out of the room when they are being talked to". Most tellingly, she found it difficult to make sense of grading-decisions even when she had a full report and list of judgement-criteria to "illuminate" them. This was suggestive of the separate and highly internal nature of the grade-decisions observers make – and the difficulty in linking these to the relatively clear identification of strengths and improvements.

#### 4.4 Summary

This chapter has presented various ‘snap-shots’ of the OTL experience: what it feels like to be observed, to carry out an observation, and to follow someone else’s thoughts as they carry out an observation. The first of these may have been familiar, in that the insights echo those of studies in the literature (O’Leary, 2013; Boocock, 2013; Cockburn, 2005). However, as these insights were personal to me, they might be justified as offering background to my positionality as researcher. The other reflections focus clearly on the experience of the observer and are, therefore, more directly relevant in setting the context for the findings to follow.

Key points:

- Being observed can be stressful for even the most experienced practitioner.
- Being observed can be time-consuming for the practitioner, with potential consequences for the quality of delivery for learners.
- Observees can decide to ‘game’ the system by planning and presenting activities, resources and styles of teaching they believe will ensure higher grades.
- OTL judgements appear to be highly context-specific.
- Observers’ OTL judgements may be affected by their relationship with and previous knowledge of their observees.
- Observer judgement may be affected by a range of factors considered ‘external’ to the teaching and learning process.
- Observers worry about the formation and delivery of their OTL decisions.
- An observer’s grade-decision can be perceived as being intuitive or ‘automatic’.
- Judgement-criteria can appear to be ‘secondary’ in nature, collected to ‘evidence’ or ‘justify’ a decision that has already been formed.

One message of this chapter has been that OTL decision-making is far from being a tidy and linear process. Judgements arise from various bases and involve considerable mitigation and internal negotiation before a final grade is arrived at. It is my belief that this chapter has set the scene effectively and raised many of the issues which will be investigated more systematically in the chapters that follow. Specifically, the reflections detailed above have particularly helped to shape my approach to the research design as a whole, and can be seen to have added contextual detail to discussions which occur later in this thesis; for example: the extent to which relationships between observer and observee can affect decision-making (see section 5.3.3, p131); the intuitive nature (see section 5.3.1, p127); and the difficulty and doubt often associated with that decision-making (see section 5.3.6, p134).



## **Chapter 5 – Findings of the Research**

### **5.1 Introduction**

This chapter presents the findings of the research. Each section brings together findings arising from multiple methods. To enhance the clarity of this section, I have decided to group the findings into four sections broadly corresponding to the areas covered by each of the research questions (RQs).

#### **Notes:**

- The subjects of the interviews (6) and questionnaires (8) form a substantial subset of the larger number of observers included in the secondary data: 14 from a total of 24.
- Quotes from interviewees are labelled from Interview A to Interview F. Quotes from respondents to questionnaires are labelled from Questionnaire 1 to Questionnaire 8.
- Most of the results cited from the interviews were substantially supported by the findings of the questionnaires. Where there was no direct corroboration, this was due to the more limited nature of the questionnaire in eliciting responses. None of the findings from interviews were contradicted by those of the questionnaires.

## **5.2 RQ1: What do observers believe about the OTL process?**

In this section, findings are derived from the observer interviews (6 subjects) and questionnaires (8 subjects).

The key findings were:

- I. All observers believed that OTL was useful and important across a range of purposes; including:
  - All identified a strong QA purpose
  - All identified a strong CPD purpose
  - Most acknowledged an inspection-focussed purpose
- II. All observers were critical of elements of the grading process
- III. All observers had concerns about the effects of institutional targets and managerial priorities on the usefulness of OTL

### **5.2.1 All observers believed that OTL was useful and important**

All the observers interviewed acknowledged the importance of OTL across a range of purposes. Fundamentally, observation allows them to “understand what happens in the classroom and the tutors to understand what happens in their classrooms” (Interview A). Generally, observers accept OTL as performing necessary purposes; they cannot know about the delivery of the service “unless we sit in the back of the classroom and find out” (Interview B).

All the observers interviewed expressed the belief that OTL was important. However, that is not to say they thought it was perfect: observations are “important, as a snapshot... [but] only a partial picture.” (Interview C).

### **5.2.2 All observers identified a QA purpose**

The interviewees believed that OTL had a dominant QA function. Learners do not directly pay for ES classes, but nevertheless the language of ‘customer-service’ was sometimes employed: OTL was carried out to ensure that learners were “getting ‘value for money’” (Interview B).

Observers accepted this as necessary, but also identified this function as a primary source of the stress associated with OTL, and experienced by observees: “it’s stressful. And they hate it” (Interview B).

OTL was seen as closely linked to managerial targets: “from the service’s point of view it is ... for checking that everything is being done the way it should” (Interview D).

This sets a barrier between the observers and the observee because, although they agreed that “you do need to know what people are doing in the classroom” (Interview E), most observers would rather concentrate on helping improve teaching quality through a more tutor-focussed approach.

### **5.2.3 All observers identified a CPD purpose**

All the observers indicated that, for them, OTL should be “a developmental thing rather than a judgement thing” (Interview C). It was thought that this CPD function

was a secondary consideration as far as Newbold's management were concerned: "the service sees it as a judgement" – but observers were "more keen to use it developmentally" (ibid).

Observers reported a tension between the CPD and QA functions, voicing reservations about the effectiveness of OTL in terms of the CPD function, seeing it as valuable for tutors only "if done properly" (Interview D), with the implication being that it often is not: "I'd like to see [OTL] be far less judgemental" (Interview A). In this context, the QA function, closely linked to the practice of grading, was seen by some as undermining the wider opportunities for CPD:

"I would hope it could be much more developmental, which can also be very affirming, but it [should not] provide a grade, as it could be a lot more experimental. If somebody is trying something out, a new way of doing something, they could ask their manager to come and observe them doing it and get constructive feedback."

(Interview E)

#### **5.2.4 Most observers identified an Ofsted inspection purpose**

Five out of six interviewees considered the relationship between OTL and Ofsted inspection significant, defining the shape of OTL within Newbold. Observations can be seen as preparation for inspection, in terms of being 'mini-Ofsted' themselves and as a tool of the QA function, to bring teaching up to inspection-readiness.

Newbold has been annually obliged to provide an internal assessment of overall teaching quality to Ofsted. This has provided justification for the use of Ofsted grades

in OTL, although there has been, in fact, no requirement that institutions have to collect their data in this way. The implication, however, has become widely believed: “Ofsted requires a grade” (Interview E). Because of this, Ofsted ‘standards’ appear to influence OTL and this was resented:

“I hated it in the past when we were told that some things would limit a grade ... It was about fashions; having to limit my grades because of what was in fashion. I suppose that came down from Ofsted.”

(Interview A)

The extent to which the adoption of Ofsted ‘standards’ was successful was disputed, partly because observers suspected that inspection judgements might be as subjective as their own: “we can say we use the same criteria that Ofsted would use - which in fact, I don’t think we do, because we don’t know what Ofsted would be looking for on any given day, let’s face it.” (Interview B)

### **5.2.5 All observers were critical of elements of the grading process**

Observers considered grading, at best, an unfortunate necessity required by management: “more for the whole service view” (Interview C); at worst, a damaging distraction from more important priorities: “the actual grading is a little bit unfair – it should be more of a CPD thing” (Interview F).

One important concern observers had about grading was the apparently arbitrary, and indeed mysterious, nature of how grades are arrived at in the first instance: “never been sure or aware of what decides a two or a one or a three” (Interview F).

Observers were aware that grading had considerable consequences for observees: “a poor observation grade is a judgement on them and they have to be re-observed” (Interview C). For this reason, observers can experience anxiety around coming to a grade decision:

“How reluctant I am to take part in the process and how hard I find it.

Because it is seen as a judgemental rather than a developmental process; it is difficult; people do feel hurt”

(Interview A)

The grade was also seen as a very limited reflection of what occurs in the classroom, simply a matter of “this is what we saw on the day”, there was a concern that “this is what the tutor will score for the year, and it just doesn’t say enough about what I know they do, the breadth of what they deliver” (Interview B).

Observers believed that there was too much focus on the grade as opposed to the other features of the observation report, such as the narrative description or the strengths and improvements. This unhelpful focus was thought to limit the extent to which guidance and feedback could lead to improved practice: “once somebody hears or sees a grade, they can just shut down” (Interview B)

#### **5.2.6 All observers had concerns about the effects of institutional targets and managerial priorities on the efficacy of OTL**

Observers believed that OTL often served the requirements of institutional targets to the detriment of its efficacy for CPD. At worst, observers feared that managerial concerns, real or perceived, might undermine the entire process; either by reducing it to

empty formality – observations “to tick a box, to give a grade” (Interview E) – or by raising the suspicions of the observees: “perceived to be used to check up on people” (Interview B).

Management was believed to use OTL as a tool for meeting its institutional targets and to respond to external pressures; for example, the requirement to constantly improve OTL grades, even when seen as unrealistic: “I’m afraid it’s getting everybody to a [grade] one” (Interview F).

### **5.2.7 Reflections on findings of RQ1**

This section illustrates the contradictory nature of OTL in the minds of the observers.

There is wide consensus as to the importance and main purposes of the process.

However, observers appear to experience an inherent tension around the simultaneous CPD and QA foci of OTL. This tension also manifests itself in observers’ criticism of applying grades; the element believed by many to be the main obstacle to OTL

becoming more usefully focussed on CPD. This reluctance to grade observations could be seen as being linked to observers’ awareness, as teachers, of the negative effects of grading learners, rather than focussing on formative feedback.

### **5.3 RQ2: How do observers explain their OTL decision-making?**

In this section, findings are derived from the observer interviews (6 subjects) and questionnaires (8 subjects).

The key findings were:

- I. All the observers interviewed said they used intuitive/holistic decision-making
- II. All the observers interviewed said they also used criterion-based decision-making
- III. All observers believed their OTL decisions were affected by their relationships with observees
- IV. All observers cited their own professional experience as a basis for decision-making
- V. All observers believe their OTL decisions are affected by external factors and considerations
- VI. All observers expressed doubts about their OTL decisions

#### **5.3.1 All observers cited intuitive/holistic decision-making**

Overwhelmingly, observers located their initial decision-making in the holistic and the intuitive: they used expressions such as “gut feeling” (Interview B); and “instinctive reaction” (Interview A).

These responses suggested that observers were basing their judgements on an internalised standard which they found difficult to explicitly rationalise. They were



happy to admit to deciding in this way, so there was evidently a general confidence that this intuitive approach had a kind of validity for them. However, it was worth noting that OTL decision-making was not considered easy, and observers worried about their judgements. The fact that they did not understand exactly how they reached their decisions served to exacerbate a sense of uncertainty that appeared to be fairly common:

“The reality is that I get a hunch, a sense of what the grade should be and then ... do I record evidence to support that hunch? I don’t know. How do we judge anybody?”

(Interview E).

Despite the awareness that their own decision-making was “largely instinctive” (Interview A); when it came to being observed themselves, observers often expressed a lack of understanding of how OTL judgements in general come into being:

“I’ve only ever had a 1 from my manager once: and I find it hard to know why, because I feel I have done better sessions since.”

(Interview A)

Observers appeared to hold both that, OTL judgement was highly personal and intuitive and, conversely, that there should be a consistently transparent rational process at work. That this appeared contradictory might explain some of the anxiety and doubt that is experienced by observers (and their observees) in relation to OTL.

In my auto-ethnographic reflection on carrying out an observation, I also experienced coming to my grade decision as an intuitive and almost automatic act (see section 4.2.2, p105).

### 5.3.2 All observers also cited criterion-based decision-making

All observers cited judgement-criteria, the lists of strengths and improvements recorded in the observation reports, as important to their decision-making. This might seem to contradict the previously stated prevalence of intuitive holistic thinking, suggesting that observers also followed a parallel 'heuristic' approach. However, the evidence suggested that observers mostly used criteria to confirm or justify judgements which have already been arrived at holistically:

“I have some sort of gut feel for the grade I think it deserves. I then go through looking for things [Judgement-Criteria] that support or contradict that instinct.”

(Questionnaire 8)

A senior manager-observer stated that he used criteria as part of a mix to “balance what I think I am seeing against the agreed criteria”, but conceded that the decision was “basically a best guess” (Questionnaire 4).

There was a lack of clarity about the role of criteria in the observers' decision-making process, and about the origin and source of those criteria

“I suppose I *have* a checklist in the back of my mind” (Interview A)

“I am *not* doing a checklist in my head” (Interview E)

It might be reasonably assumed that there is a standard set of criteria: “I always assumed they came from Ofsted; the criteria that Ofsted uses” (interview F). However, in reality, although there have been attempts to introduce standardised criteria, to date none have been universally accepted within Newbold: “they used to have all sorts of

standards, they kept changing them, which I used to try and follow, but now I don't”  
(Interview A).

Judgement-criteria were informally held and had been accumulated from a variety of sources: Ofsted, training, management priorities, and common-sense. However, it was clear that these were not necessarily applied in a checklist fashion, again still favouring a holistic approach:

“I don't go through criteria by criteria, have they done this? Have they done that? I get a sense of their overall ability”

(Interview E)

One method of applying judgement-criteria was to privilege certain criteria over others, making them in effect *key decision factors*. The judgement-criterion most often cited as decisive was ‘Evidence of Learning’:

“There is one thing I am looking for really: are the learners learning?  
Everything else feeds into that. I have strengths and weaknesses, but it is all  
around ‘are learners learning?’”

(Interview A)

Some observers were able to explicitly state how they used this criterion as a limiting factor: “If no learning is witnessed then the grade would be 4” (Questionnaire 6).

Another oft-cited method was to consider the numerical balance of strengths and improvements: “I think I would weigh the things I was looking at. There might be four good items; four unsatisfactory” (Interview D).

However, there was always, for some, the belief that the concept of ‘effective teaching’ might exist quite separately from any of the criteria used to define it, or purport to comprise it:

“I am quite convinced that you can teach brilliantly and get your learners to learn without doing any of those things in the checklist in the back of my head.”

(Interview A)

This commonly felt, almost romantic, appeal to the ineffability of what goes into making ‘effective teaching’, might indicate the existence of an underlying experience-based Wisdom of Practice at the heart of practitioners’ understanding of their work.

### **5.3.3 All observers believed their OTL decisions were affected by their relationships with observees**

Observers all acknowledged that their relationships with their observees could have a profound effect on their decision-making: “It is about everything we know about that tutor” including potentially “their past history” (Interview B). At Newbold, most observations were carried out by the observee’s line manager – a colleague they may have known for years, and been observed by numerous times before. This situation made in-house OTL very different from the model of the Ofsted inspection – in which observees are unknown to observers. This appeared to have consequences for observer decision-making:

“One of the things that I find so difficult is that I have been managing the same group of people for so long now. For about 10 years, I’ve been observing the same people doing the same sorts of things ...”

(Interview A)

This situation raised questions of fairness and objectivity for some of the observers: “we do have some idea of what to expect, what a tutor is capable of. However, I am fair and try to only report/grade what I see during the observation” (Questionnaire 5). Some observers tried to be “as impartial” as possible; not allowing “grading to be influenced by previous observations/knowledge of a tutor” (Questionnaire 6). Other observers, however, felt that prior knowledge was a positive benefit in reaching an accurate assessment: “the core of their teaching is going to be the same ... when you’ve observed someone a lot, over a period of time; you get to have a feeling for that” (Interview E). Prior experience of the observee allowed these observers to overlook a momentary lapse, to get closer to the ‘essence’ of the observee’s ability: “I do know that this tutor does rack up the pace, normally. They were having a bad day...” (Interview D).

Observers identified another benefit of their relationships with the observees: the ability to take the specific needs of the individual into consideration. This was seen as a question of sensitivity to needs, a key concern for observers:

“We can’t treat all tutors as the same, everybody is an individual, some are more robust than others and can take the criticism others need to be treated more gently – not because they are not good teachers ... but because they need a bit more consideration when it comes to putting them in the observation process.”  
(Interview D)

The issue of the effect of the relationship between observer and observee was also clearly illustrated in the ethnographic reflection on co-observing with an experienced observer (see section 4.3.1, p111).

#### **5.3.4 All observers cited their own professional experience as a basis for decision-making**

Observers located a significant part of their decision-making in their prior professional experience, both as observers – “weighted with the experience over many years of observations” (Questionnaire 4) – and as practitioners themselves: “I suppose a lot of it also has to be subjective – based on your knowledge of teaching” (Interview F).

This accrual of experience was held to provide a form of internalised standard, against which judgements were formed: “I have an idea in my mind of what a ... good lesson is – I measure [the observed lesson] against that” (Interview C).

Ultimately, this appeal to experience at the core of the judgement process, and its internalised nature, begins to suggest what has been discussed elsewhere as *Wisdom of Practice*: a set of knowledge so familiar and seasoned in the mind of the practitioner that it reaches a level of automaticity that is difficult to explicitly rationalise:

“I think that, after a while, it’s like driving a car – it’s just there, you don’t have to think about it. It’s not at the forefront of my mind; it’s there in the background.”

(Interview D)

#### **5.3.5 All observers believed their OTL decisions were affected by external factors and considerations**

During the course of the interviews and questionnaires, observers cited various external factors they believed might impinge on their decision-making process; for example:

- Availability of volunteer support
- Learner behaviour/personal issues
- Tutor personal/family issues
- Technological failures
- Punctuality and attendance of learners

There was considerable variation between the observers as to what constituted an external factor, as opposed to something which was within the control of the observee. ‘Attendance’ of learners was sometimes present as a judgement-criterion in the observation reports, but there was also a difference of opinion about whether such an ‘external factor’ should be ignored or be accounted for in the grade judgement: “I don’t think that tutors should be blamed for attendance or punctuality” (Interview D).

How a tutor handles an external factor may also provide an opportunity to demonstrate higher ability: “how they cope with anything that isn’t controllable ... the flexibility of the tutor to deal with unforeseen circumstances” (Interview F). This raises the question of how the observee’s performance would have been rated if such an unexpected external factor *had not* happened to arise – would lessons that go to plan be marked down?

### **5.3.6 All observers expressed doubts about their OTL decisions**

Even an observer with decades of experience was able to identify moments within their current OTL practice when doubt got the better of them:

“Although I’ve been doing this since Noah was a lad, there are still days when I go in and I think, I don’t know what to do with that. It may be that I’m having a bad day.”

(Interview B)

The prevalence of doubt expressed by observers might be understood as a consequence of the apparent ‘ineffability’ of OTL decision-making: “to be honest, I find the whole process a bit incoherent” (Interview C). Perhaps surprisingly, even amongst a group of experienced observers, there persists the sense of fundamental mystery:

“I would love there to be some more formulaic approach, but I suspect it is impossible. There are too many variables, too many different approaches and too many classes with widely differing needs.”

(Questionnaire 8)

Observers’ found it difficult to adequately explain where their OTL judgements come from: “obviously the feeling comes from somewhere...” (Questionnaire 7).

A key aspect of OTL, where doubts commonly arise, was the grade-decision. Some observers reported finding it difficult to reach a decision particularly with such high stakes for the observee:

“There was an observation that I did and I spent hours [internally] wrangling over it – which was not a good position to be in.”

(Interview B)

There was also the issue of their professional standing as observers: “will my judgment be questioned? Who is to say I am right?” (Interview B). Deciding at grade boundaries



was often experienced as the most problematic, “agonising” (Questionnaire 8), part of the decision process, and one that observers saw as unnecessarily difficult:

“It’s a spectrum and where the arbitrary line between 3 and 2 is, is always a bit tricky”

(Questionnaire 7)

The word “arbitrary” in this quote underlined the curious sense of ineffability within the OTL system. Most observers considered the Grade 2 / 3 boundary the most problematic – understandably, because the decision to cast an observee into the ‘unsatisfactory’ category carried considerable professional and emotional consequences. However, observers reported difficulty with grade-decisions at all levels - “I stress most about whether a tutor should be given a Grade 1 or 2” (Interview E) – and this suggested that the difficulty might also be about comprehending the differences between levels:

“No one has really demonstrated how wide that boundary is – Grade 2 is so broad, surely Grade 1 must also have some ‘width’ – what do we use to gauge that?”

(Interview B)

The ethnographic reflections based on experiences of co-observing also illustrated this area of doubt in operation (see sections 4.3.1 p111 & 4.3.2 p113)

### **5.3.7 Reflections on the findings of RQ2:**

It appeared that observers were not readily able to express where their OTL decisions come from. The apparent contradiction at the heart of these findings was that all the

observers believed both that they use a “gut-feeling”, holistic approach to decision-making and, simultaneously, a more systematic criteria-based process. Clearly, this cannot be literally the case – and this speaks to the considerable anxiety that many observers report concerning their decision-making. What emerges is a sense that the OTL decision is a two-stage affair: with observers forming initial judgements based on deep experience-based knowledge and understanding (potentially the ‘Wisdom of Practice’), but then refining and/or confirming these judgements using what they see as the criteria of ‘effective teaching’, from whatever source they may be drawn.

The extent to which the personal relationship between observer and observee might play a part in the OTL decision was also a significant finding.

#### **5.4 RQ3: What elements of ‘effective teaching’ do observers focus on in evidencing their OTL decisions?**

In this section, findings are derived from the interviews (6 subjects) and questionnaires (8 subjects) of observer-managers, and the analysis of secondary data (24 subjects inclusive of the aforementioned).

The key findings were:

- I. All the observers questioned cited judgement-criteria as the main evidence for their OTL decisions
- II. It was possible to collate and rank the most frequently occurring judgement-criteria in the observation reports
- III. Evidence from the data suggests that observers operate a ‘hierarchy of criteria’ in which some judgement-criteria are valued less highly than others.

- IV. There was a clear relationship between OTL grade and the incidence of frequently cited judgement-criteria as strengths and improvements
- V. Each OTL grade is associated with a different profile of specific judgement-criteria

Note: judgement-criteria referred to here - originating from comments made by observers, verbally or as judgements in the observation reports – were originally expressed in various ways, and have been clarified for ease of handling (as discussed in section 3.5.4, p81).

#### **5.4.1 All observers questioned cited judgement-criteria as the main evidence for their OTL decisions**

When asked what they were looking for in an observed session, observers began to list elements of teaching closely resembling the judgement-criteria of the observation reports. These elements appeared to emerge from the experience, knowledge and practice of the observers: “I am aware of certain things that I believe make a ‘good’ teacher” (Questionnaire 4).

There have been attempts within Newbold to create a standardised list of judgement-criteria, but these seem to fall by the wayside of observer practice, some observers admitted that they “don’t even know if such a thing exists” (Interview D).

Although the elements of ‘effective teaching’ mentioned in interviews and questionnaires consistently reflected those appearing as strengths and improvements in observation reports (see section 5.4.2, below), there were examples of inconsistency in the priorities expressed by the observers and the reality of the evidence in the secondary

data. The notable example is of the criterion ‘Learner Voice’: most observers interviewed considered ‘learner voice’ an important factor in their decision-making, although it only occurred as a judgement-criterion in 9% of observations.

#### 5.4.2 Most frequently occurring judgement-criteria in the observation reports

The observation reports were compiled by the full complement of 24 ES observer-managers, which includes the 14 subjects of the interviews and questionnaires.

**Table 5.1**, below, shows the top twenty judgement-criteria cited in the observation reports by frequency for both strengths and improvements (for a complete list of the judgement-criteria recorded in the observation reports see **Appendix 1**, p217).

Judgement-criteria	Overall		As Strengths		As improvements	
	frq	% occ	frq	% occ	frq	% occ
Differentiation	88	64.23	48	35.04	40	29.2
Feedback	70	51.09	38	27.74	32	23.36
Planning	70	51.09	47	34.31	23	16.79
ICT	65	47.45	42	30.66	23	16.79
Activity	63	45.99	52	37.96	11	8.03
Questioning technique	63	45.99	41	29.93	22	16.06
Classroom management	60	43.8	44	32.12	16	11.68
Pace	49	35.77	21	15.33	28	20.44
Resources	49	35.77	46	33.58	3	2.19
Teaching methods	49	35.77	16	11.68	33	24.09
Subject content	46	33.58	32	23.36	14	10.22
Aims	45	32.85	20	14.6	25	18.25
Peer working	39	28.47	25	18.25	14	10.22
Reflective practice	36	26.28	19	13.87	17	12.41
Paperwork	33	24.09	22	16.06	11	8.03
Accreditation focus	31	22.63	31	22.63		
Clear instructions	31	22.63	8	5.84	23	16.79
EandD	31	22.63	25	18.25	6	4.38
Assessment	30	21.9	20	14.6	10	7.3
Engagement	29	21.17	27	19.71	2	1.46

*Table 5.1: Top 20 OTL judgement-criteria cited in observation reports overall by frequency (frq)*

By comparing the ten most frequently occurring strengths and the ten most frequently occurring improvements, a set of seven key judgement-criteria emerged which occurred within the top ten of both strengths and improvements (shaded in pink).

These seven criteria might be considered to constitute the core criteria across the range of the sample:

- Differentiation
- Feedback
- Planning
- ICT
- Questioning Technique
- Pace
- Teaching Methods

Explanations of less familiar criteria are given in **Chapter 2 (Table 2.6 p54)**, here I comment on some trends within the data:

Judgement Criterion	Commentary
Differentiation	<ul style="list-style-type: none"> <li>• The most commonly cited criterion</li> <li>• Considered by observers a key skill in “meeting learners’ needs” (Questionnaire 5).</li> <li>• For some, at the heart of the learner-centred approach.</li> </ul>

Feedback	<ul style="list-style-type: none"> <li>• A key element of Assessment for Learning (AfL)</li> <li>• Promoted to observers at training sessions.</li> <li>• Only mentioned once by an observer in interview or questionnaire.</li> </ul>
Planning	<ul style="list-style-type: none"> <li>• Almost twice as often cited as strength than improvement.</li> </ul>
ICT	<ul style="list-style-type: none"> <li>• Much more likely to figure as a strength</li> <li>• Has previously operated as a management-imposed “limiting factor” (Interview A), which may help to explain its continued prominence.</li> </ul>
Activity	<ul style="list-style-type: none"> <li>• Much more likely to figure as a strength</li> <li>• More often cited as a strength in lower graded observation reports</li> <li>• Seldom mentioned at the highest level.</li> </ul>
Questioning technique	<ul style="list-style-type: none"> <li>• More likely to figure as a strength</li> </ul>
Classroom management	<ul style="list-style-type: none"> <li>• Much more likely to figure as a strength</li> </ul>
Pace	<ul style="list-style-type: none"> <li>• More often cited as an improvement</li> </ul>
Resources	<ul style="list-style-type: none"> <li>• The most strongly differentiated criterion: more often a strength by 32 percentage points</li> <li>• Figures as an improvement in only 3 observation reports.</li> </ul>

Teaching methods	<ul style="list-style-type: none"> <li>• Twice as often cited as an improvement than as a strength</li> </ul>
Clear instructions	<ul style="list-style-type: none"> <li>• Nearly three times more often cited as an improvement (17%) than a strength (6%).</li> </ul>
Accreditation focus	<ul style="list-style-type: none"> <li>• Perhaps unexpectedly, given the funding priorities of Newbold, this criterion does not figure in a single observation report as an improvement</li> </ul>

*Table 5.2: Judgement-criteria and some trends in the data*

### **5.4.3 Evidence from the data suggested that observers operated a ‘hierarchy of criteria’ in which some judgement-criteria are valued less highly than others**

From **Table 5.3** (p144), it could be seen that judgement-criteria profiles vary considerably between the grades: certain criteria appear more closely related to higher grades than others. This supported the finding that observers operated a kind of ‘hierarchy of skills’ with more basic elements recorded at the lower grades but taken for granted in higher graded lessons. This was most clearly demonstrated in the case of ‘differentiation’: cited as an area for improvement in 53% of lower graded observations, but not at all in the highest graded sessions.

Most of the observers interviewed acknowledged the existence of a hierarchy of judgement-criteria. They would “weight differently, different aspects of the observation process” (Interview D). There was an acknowledgement that there were “some key things that simply do make a difference” (Interview B) and others that did not. All observers cited ‘Evidence of learning’ as a highly privileged factor in decision-making,

for example; whilst other criteria, notably 'Paperwork', were often thought of as lacking relevance:

"If the unsatisfactory are things like paperwork, I would probably come down on the side of good. But if the paperwork was all spotless and fantastic but actually their teaching methods or their interaction with the learners or some aspect of knowledge or pace – I would probably come down on the side of a three."

(Interview D)

It should be noted that, despite such low regard for 'Paperwork' as a judgement-criterion among observers, it was cited in 24% of the observation reports, most often as a strength. This high occurrence might have reflected then current managerial prioritisation, or perhaps be accounted for by the relatively easy and obvious nature of the criterion's availability for assessment: paperwork being either present-and-correct, or not.

#### **5.4.4 There was a clear relationship between OTL grade and incidence of frequently cited judgement-criteria as strengths and improvements**

**Table 5.3**, (p144) shows the percentage frequency of occurrence of the ten most common judgement-criteria in the observation reports for each grade, as both strengths and improvements.

Based on the variation of frequency of each judgement-criterion when considered separately by grade, it was clear that almost all of the criteria showed a pattern of related progression: frequency of strengths rising with increasing grades, and frequency of



improvements increasing, and vice versa. This tendency might appear to confirm a relationship between the frequency of the judgement-criteria and the grades. Overall, this result suggests that judgement-criteria were significant in relation to the observation grade.

An exception here was ‘Activity’, which did not appear as frequently as might be expected at Grade 1 – perhaps because this strength was taken for granted at the very highest level. ‘Feedback’ also occurred as a strength more frequently at Grades 3/4 than at Grade 2 – perhaps due to the concentration of CPD on this skill that had taken place during the months leading up to this research and which may have been targeted towards the lower achieving observees in particular.

Judgement-criteria	%age occurrence by OTL Grade					
	Grade 1		Grade 2		Grade 3/4	
	strengths	imps	strengths	imps	strengths	imps
Differentiation	67%	0%	42%	19%	14%	53%
Feedback	44%	6%	23%	19%	30%	35%
Planning	39%	0%	36%	13%	26%	33%
ICT	50%	6%	36%	13%	12%	28%
Activity	22%	0%	49%	6%	23%	14%
Questioning technique	50%	0%	31%	17%	19%	21%
Classroom management	56%	0%	34%	12%	16%	19%
Pace	22%	0%	19%	21%	5%	28%
Resources	44%	0%	36%	3%	21%	5%
Teaching methods	22%	0%	14%	13%	5%	51%

Table 5.3: Frequency of judgements by grade (as strengths and improvements)

#### 5.4.5 Each OTL grade was associated with a different profile of specific judgement-criteria

When lists of most frequently occurring criteria were grouped by OTL grade (see **Table 5.4**, below), what emerged was a separate profile of judgement-criteria associated with

particular grades - each with its own characteristic range of skills, abilities and relative weaknesses.

Grade 1		Grade 2		Grade 3/4	
Strengths	imps	Strengths	imps	Strengths	imps
Differentiation	Aims*	Activity	Pace	Feedback	Differentiation
Classroom management	Reflective practice*	Differentiation	Feedback	Subject content	Teaching methods
ICT	Feedback*	ICT	Differentiation	Planning	Feedback
Questioning technique	ICT*	Resources	Questioning technique	Activity	Clear instructions
Feedback		Planning	ICT	Resources	Planning

*Table 5.4: Most frequent 5 strengths and improvements (imps) by grades (\*occurring in 6% or fewer cases)*

These profiles were, of course, based on amalgamated data, but for illustrative purposes it might be useful to examine each one as if they were a ‘typical’ example of a practitioner at the corresponding level of perceived teacher effectiveness:

### **The Grade 1 Tutor – the outstanding performer**

- This practitioner was aware, and in control, of the various competing aspects of a mixed-ability learning environment, as demonstrated by outstanding skill in ‘Differentiation.
- Particularly strong on the interpersonal skills involved in successful classroom management; handling discussion and question and answer sessions very effectively.

- The Grade 1 Tutor was well versed in the feedback and questioning techniques associated with Assessment for Learning, and was able to apply them to effect in the classroom.
- The Grade 1 Tutor seldom displayed any noticeable areas for improvement, but those mentioned tended to be low in the 'hierarchy': paperwork (writing up reflective diaries) or routine practice (displaying aims at the start of the lesson), for example.

### **The Grade 2 Tutor – the standard practitioner**

- The basics of classroom practice were safe in this practitioner's hands: planning lessons and activities, and designing effective resources.
- The Grade 2 Tutor was often able to differentiate work for individual learner needs, although not always most efficiently.
- Because of an ability to plan purposeful learning sessions, the lessons may be effective, rather than inspired: a key criticism might be a lack of pace.
- The Grade 2 tutor was also less likely than the Grade 1 Tutor, to make effective use of the Assessment for Learning (AfL) skills of formative feedback and questioning techniques; and it was in these areas that there was room for improvement through CPD.

### **The Grade 3/4 Tutor – the poor communicator**

- This practitioner lacked ability to effectively communicate their knowledge of subject.
- The Grade 3/4 Tutor demonstrated a narrower range of teaching techniques, and there may have been a tendency for the sessions to become tutor-led.
- Despite often being able to plan good individual activities and effective resources, the Grade 3/4 Tutor might struggle to structure these into an overall lesson plan or to tailor material and approach appropriately for all the learners' differing needs.
- Crucially, the Grade 3/4 Tutor appeared less confident with the interpersonal skills involved in offering clear instructions and explanations in class, and in giving appropriate feedback to learners.

#### **5.4.6 Reflection on the findings of RQ3**

Individual criteria appeared to be linked differentially to OTL decision-making; with the most frequently cited judgement-criteria strongly related to grading. This appeared to support the belief, held by most observers, that there was a 'hierarchy' of criteria - with some elements of 'effective teaching' more significant than others.

Individual profiles of judgement-criteria emerged related to each OTL grade and their comparison may be of some illustrative value, perhaps informing the focus of CPD guidance and other training.

## **5.5 RQ4: How do observers' judgements relate to other potential outcome measures of 'effective teaching'?**

In this section, findings are derived from the observer interviews (6 subjects) and questionnaires (8 subjects), and the analysis of secondary data (24 subjects inclusive of the aforementioned).

The key findings were:

- I. All observers questioned considered accreditation an important OTL focus - although most reported unease about the relationship between OTL and accreditation outcomes
- II. Achievement and accreditation rates had no observable relationship with OTL grades *overall* (but a difference was observed at the boundary of the acceptable OTL grades)
- III. Attendance rates had no observable relationship with OTL grades
- IV. Separate outcome measures were associated with different profiles of specific judgement-criteria

### **5.5.1 All observers questioned considered accreditation an important OTL focus**

All observers interviewed acknowledged the importance of learner achievement and accreditation - "now more than ever it is about getting qualifications" - even if it was for reasons external to the classroom: "that is where we get our funding" (Interview D).

However, in their interviews, a majority of observers also expressed a sense of unease about the relationship between OTL and accreditation: "You can't just judge their practice by whether they get people through the odd [qualification]" (Interview E).

Observers acknowledged, but found it hard to explain, the fact that the connection between their ideas of ‘effective teaching’ - as represented by a high OTL grade - and levels of accreditation and achievement did not appear straightforward:

“A poor observation grade is a judgement on [the observee] and they have to be re-observed, but if they are getting good results – then something is working that we haven’t seen.”

(Interview C)

Whatever this “something” was, it was not *necessarily* seen as an indicator of ‘effective teaching’:

“It might be that their approach is purely exam focussed which might give us a false success in some ways ... they might not be giving them the broader skills we’d also want to include for the community”

(Interview C)

It should be noted that, although considered an important factor, analysis of the secondary data showed that ‘Accreditation Focus’ appeared as a judgement-criterion in less than a quarter of the observation reports (23%). This perhaps indicated that, while accreditation might have been a current managerial priority at Newbold, its presence, or absence, as a focus in the classroom was not of the first order of importance for observers when considering ‘effective teaching’ - where the highest grades were more often associated with a set of qualities some referred to as “buzz” (Questionnaire 3):

“There is that ‘X-Factor’ thing about it ... there is something about tutors who are naturally outgoing and creative in their voice and in their language that does just give that little bit extra”

(Interview B)

### 5.5.2 Achievement and accreditation rates had no observable relationship with OTL grades *overall* (but a difference was observed at the boundary of the acceptable OTL grades)

**Table 5.6** compares observees' average OTL grades (2012-2014) with the same observees' average accreditation and achievement rates, for the same period. It could clearly be seen that there was little difference in terms of achievement and accreditation rates when compared across the full range of OTL grades. This would suggest a lack of any obvious relationship.

Grade Average	Average Accreditation Rates	Average Achievement Rates
1 to < 1.5	59	89
1.5 to < 2	54	88
2 to < 2.5	57	90
2.5 to < 3	54	91
3 to $\leq$ 4	46	78

*Table 5.5: Average OTL grades against average accreditation and achievement rates*

However, the results do show a marked difference at the boundary between Grades 2.5 to < 3 and Grades 3 to  $\leq$ 4 - a gap of 8 and 13 in accreditation and achievement rates respectively. This is a clear 'break' between the 'acceptable' grades (1 and 2) and those constituting an 'unsatisfactory' result (3 and 4).

It was, perhaps, predictable that those observees whose teaching was considered the least satisfactory would also be the ones whose learners showed the least success in terms of achievement. However, this finding might have been expected to be replicated smoothly across the full grade range – and this was not the case. It appears that progressively higher levels of teacher effectiveness, as measured by grades, do not result in higher levels of learner achievement, as measured by results.

### **5.5.3 Attendance rates had no observable relationship with OTL grades**

**Table 5.6**, below, shows a comparison between observees' average OTL grades (see note on numerical representation, p79) with the same observees' average attendance rates. Again, as with the achievement and accreditation outcomes, it could easily be seen that there was little association between OTL grade and attendance rates. In this case, the lack of differentiation was constant across the whole range of grades. In fact, it can be seen that the highest average attendance was sitting in the exact mid-range of the OTL scores. It was also the case that, although the data were very bunched, the lowest average attendance rate was recorded for observees with the highest OTL grades. A possible explanation for this being that perhaps the more proficient teachers may be the ones more likely to say to unsuitable learners, "this class is not for you."



<b>Grade Average</b>	<b>Average Attendance Rates</b>
1 to < 1.5	70
1.5 to < 2	72
2 to < 2.5	75
2.5 to < 3	71
3 to ≤4	73

*Table 5.6: Average OTL grades compared with average attendance rates*

It seemed reasonable to conclude from this analysis that the effectiveness of teaching, as recorded by OTL, had no relationship with learner attendance rates.

‘Attendance’ also existed as a judgement-criterion in the observation report data.

However, this was low in frequency, featuring in less than 12% of observation reports (10 mentions as strength; 6 mentions as improvement). Only one observer mentioned attendance in interview, and this was to complain that the criterion was inappropriate as a measure of ‘effective teaching’:

“I don’t think that tutors should be blamed for attendance or punctuality.

There are so many factors that affect this... ”

(Interview D)

#### 5.5.4 Separate outcomes were associated with different profiles of specific judgement-criteria

**Table 5.7**, below, shows a comparison of the frequency of occurrence of OTL judgement-criteria for the 20% of observees achieving the highest rates of accreditation and attendance. It can be seen that a different profile of judgement-criteria as strengths emerged in each case, demonstrating that different sets of skills and abilities might be considered more relevant to each outcome measure.

%age frequency in observation reports highest accreditation rate tutors		%age frequency in observation reports highest attendance rate tutors	
Planning	53	Differentiation	54
Resources	41	Classroom management	46
Differentiation	38	Activity	38
Subject content	32	Enthusiastic delivery	38
Engagement	29	Evidence of learning	38
Equality and Diversity	24	Planning	38

*Table 5.7: Profiles of judgement-criteria as strengths for high accreditation rate, attendance rate and grade 1 observations (for comparison)*

Although, there was some overlap in criteria between the two lists, albeit with differing priorities, it was apparent that classroom practitioners achieving the higher rates of either outcome measure were associated with different patterns of skills and abilities:

### **The High Accreditation-rate Tutor**

- This practitioner combined attention to the basics of teaching – lesson planning and preparation of resources are particular strengths – with a good knowledge of the learners and their individual needs ('Differentiation').
- This understanding of the learners extended to an awareness of, and sensitivity to their differing backgrounds, cultures and 'able-ness' ('Equality and Diversity').
- This teacher's knowledge of subject was very evident and brought to bear effectively in the lesson design and presentation.
- The learners were engaged by this tutor's teaching, perhaps because what was being delivered was clearly relevant to them in terms of their accreditation goals.

### **The High Attendance-rate Tutor**

- This tutor was keenly aware of the learners' individual learning needs and effectively tailored activities to deliver personalised learning ('Differentiation').
- The quality of classroom management, combined with the teacher's enthusiastic delivery, may have been central to explaining why the learners were more likely to attend.
- Lessons were well planned and there was real sense of learning taking place in this tutor's classroom.

It was interesting to note that the profile of judgement-criteria for the 'Grade 1 Tutor' (see **Table 5.4**, p145) was more similar to that of the 'High Attendance-rate Tutor' than to that of the 'High Accreditation-rate Tutor'. Whilst this may seem counterintuitive – we might expect the highest quality teachers to be more like those achieving the highest

rates of accreditation – it might be explained by the fact that preparation for examinations is much more of a ‘one-size-fits-all’ type of exercise, with work content dictated by the exigencies of the test rather than the wider needs of the learner. Whereas, receiving tuition and resources that seem tailor-made to fit their own needs and sessions that are enthusiastically delivered would be much more comfortable and attractive for learners – making them more likely to continue attending the course.

#### **5.5.5 Reflection on findings of RQ4:**

It was clear from these findings that any putative connection between ‘effective teaching’ and other outcomes was far from straightforward. The common-sense expectation that high levels of attendance might correlate with the higher OTL grades is not supported here, despite the apparent overlap in profile of judgement-criteria with that of the Grade 1 tutor.

That levels of accreditation and achievement rate only demonstrated a noticeable difference at the boundaries between ‘satisfactory’ and ‘unsatisfactory’ grades, and failed to correlate across the full range of grades, might indicate the lack of effectiveness of OTL grading – if, as some sources suggest, such outcomes were the truest measure of ‘effective teaching’. However, the ability of the OTL process to identify the least effective practitioners in terms of this important funding-related outcome, might suggest some the value of OTL as a QA tool in questions of basic competency.

It was clear from the separate profiles of judgement-criteria that emerged for observees achieving different grades, and the highest rates accreditation and attendance, that unique combinations of skills and aspects of teaching appeared relevant to different

outcomes. An awareness of this range of differences might be usefully applied to the analysis of the results of OTL in a CPD context.

## **5.6 Summary of key findings**

Here, I shall briefly mention the findings I consider the most significant in answering my overarching question: “How do observers make judgements in OTL?”

**All observers were critical of the grading and had concerns about the intrusion of institutional targets and managerial priorities on the usefulness of OTL.**

This might indicate that observers believe the judgements they want to make about ‘effective teaching’ might be undermined by the need to match those judgements to the requirements of grading, and that targets and priorities they consider extraneous to the field of ‘effective teaching’ are resented.

**Although apparently contradictory, observers appeared to use both intuitive/holistic and criteria-based decision-making methods.**

The apparent contradiction might be considered to indicate a dual process at work. This is a key finding because it demonstrates that observers rely on an ‘automatic’ process to reach their decision about the teaching they have observed, and then use a criteria-based method to both refine and justify that decision.

**Observers revealed that their OTL decisions were affected by their relationships with observees, and cited their own professional experience as a basis for decision-making.**

These findings confirm other aspects of the observers' decision-making process: factors external to the actual observed lesson played a part in their deliberations, and that they are aware that their own experience as both observers and as practitioners is an important basis for their judgements. This latter point also begins to support the role of a Wisdom of Practice.

**All observers expressed doubts about their OTL decisions.**

Whilst it may be understandable that committing to high-stakes decisions may be a source of anxiety, it is a notable feature of the OTL decision-making process that the observers often doubt the validity of the judgements they arrive at. This may be an indication of their unease with the grading process. However, there was also an element of doubt associated with the expressed fact that observers often do not understand how they made their decision – another indication that their decision-making had become largely automatic and based on a pool of experience and knowledge such as the Wisdom of Practice.

**Achievement and accreditation rates had no observable relationship with OTL grades overall; attendance rates had no observable relationship with OTL grades**

The nature of observers' decision-making appears, in the light of these findings, to centre on a different set of priorities than that associated with either success in accreditation or the ability to keep learners attending classes regularly. This insight has implications for the current debate surrounding additional or alternative outcome measures of 'effective teaching'.

## **Chapter 6 – Discussion**

### **6.1 Introduction**

In this chapter, I discuss my findings, and the wider implications that might be drawn from them. I set out to examine how observer-managers reached their OTL judgements; what those judgements consisted of; and the extent to which the concept of Wisdom of Practice might be said to play a part in that decision-making process. As there was very little literature specifically dealing with how OTL decision-making was achieved, these findings can be considered a contribution to knowledge.

This chapter is divided into three sections:

The first section will consider the findings, organised according to the four research questions that have provided focus throughout:

- I. What do observers believe about the OTL process?
- II. How do observers explain their OTL decision-making?
- III. What elements of ‘effective teaching’ do observers focus on in evidencing their OTL decisions?
- IV. How do observers’ judgements relate to other potential outcome measures of ‘effective teaching’?

The consistency of the findings from different methodologies is also discussed.

The second section considers the relationship of OTL decision-making to the concept of Wisdom of Practice.

Finally, in the third section, I present, and critically assess, a proposed theoretical model of the OTL judgement process; possibly the first model offered that attempts to explain how observers reach their decisions.

## **6.2 Discussion of findings by research question**

### **6.2.1 What do observers believe about the OTL process?**

From an outside perspective, it might seem unsurprising that experienced observers believed in the importance and usefulness of OTL; it was a prominent part of their jobs and, in this, they appeared to support the emphasis placed on OTL by both senior management and government. This was despite the difficulty observers said they experienced in reaching their grade-decisions. The reasons for this difficulty ranged from their awareness of the emotional and professional impact they knew the decision might have on observees, to their apparent uncertainty as to how they reached their OTL decisions in the first place.

As found in the literature (Wragg, 1999; Gosling, 2002), observers were aware of the multiple purposes of OTL, especially the QA and CPD functions, and experienced a tension between these purposes. Whilst acknowledging the duty incumbent on them to ensure the provision of a high quality service for learners and other stake-holders, observers expressed a clear preference for the CPD aspects of the OTL process. They saw the demands of the QA function as getting in the way of the developmental opportunities they would ideally prioritise, confirming findings in the literature (O’Leary, 2013; Cockburn; 2005). This was a source of considerable frustration.

Confirming O’Leary’s (2013) finding that OTL was often used as a preparation for external inspection, a “mock Ofsted” (p708), observers acknowledged the influence of Ofsted on their own conduct and understanding of OTL. Observers saw the requirements of inspection as both a constraint on their CPD focus, and as a defining factor within OTL: the inspectors’ supposed standards and criteria having a major influence on their own, at least notionally.



For observers, the most contentious aspect of the OTL system was grading. Observers identified grading as having almost entirely negative effects on both observees and the usefulness of OTL itself. Grading was felt to overshadow the developmental opportunities of OTL, becoming a source of anxiety for the observee. These concerns reflected the views expressed by practitioners in various studies in the literature (Cockburn (2005), O’Leary (2006), Boocock (2013)); what is underlined here is the extent to which the observers were aware of the stress caused to their observees, and the fact that this awareness also had implications for their decision-making

Related to their concerns about grading, and in particular their sense that grading was mainly an institutional priority, observers also believed that the positive possibilities of OTL were adversely affected by institutional targets and managerial priorities.

Observers believed that it was these priorities that drove the focus of OTL from their preferred CPD to the more judgemental QA purpose. In this respect, the findings echoed those of studies that critiqued OTL as a managerial, or even neo-Fordian, device (Boocock, 2013; Cockburn, 2005).

### **6.2.2 How do observers explain their OTL decision-making?**

Whilst experienced observers were clear that they knew ‘effective teaching’ when they saw it, they found it hard to express how these judgements were made and were also unsure and doubtful about them – perhaps because they held an inconsistent dual explanation of the decision-making process. All of the observers believed that they made holistic judgements: a ‘gut-feeling’, a ‘hunch’, in line with the literature on Ofsted inspectors (Wilcox and Gray, 1996). Apparently simultaneously, observers also said

they based their decisions upon the consideration of a set of judgement-criteria; a heuristic approach that had the appearance of a more 'rational' structure.

In fact, the final OTL grade judgement appeared to be a product of the initial 'gut' reaction and a subsequent refinement process which included the judgement-criteria being applied alongside other localised considerations: such as, management priorities, paradigmatic shifts in inspection frameworks, and other external factors. This combination of the holistic and the heuristic is anticipated in the literature (Hastie and Dawes, 2010).

An important finding of this research is that observers admitted that their decision-making could be affected by their relationship with the observee. Although never appearing on the official observation report, and therefore not an explicit factor, nonetheless, personal issues played a part in the grade-decision. Very few of the observers, however, were concerned that these interpersonal aspects might be problematic. Rather, they saw considerations of this kind as playing a part in arriving at 'fair' judgements that reflected the lived reality of the context.

Observers were clear that professional experience was the key element in their decision-making. They referred to the accrual of internalised standards, based on their lived practice, against which they could form their judgements. This access to experiential data was explicitly linked with the automatic nature of the observers' holistic decision-making, recalling aspects of expert judgement referred to in the literature (Brown and McIntyre 1993, Hastie and Dawes, 2010).

A more conscious appeal to experience was made in the observers' acknowledgement of the role of comparison. This was related to the observers' knowledge of and relationships with their observees, and was based on comparing the 'quality' of one

observee's performance with another's; or one instance of an observee's practice with a previous observation. This finding introduced the idea that observers operated an informal norm-referenced scale, alongside their holistic and criterion-based methods, against which they ranked examples of practice. A problem with reliance on this method, and one that some observers acknowledged, was that comparison of this sort might lead to self-fulfilling judgements. Some observers make a conscious effort to avoid being influenced in this way; whilst others positively endorse it as a useful method.

No matter how experienced, all observers expressed doubts about their decision-making; they worried about both the accuracy of the judgements and their potentially high-stakes consequences. Much of this doubt appeared to be located in the essential ineffability of the OTL decision-making process; observers were able come to decisions, but not readily able to express why. This was particularly true of the grade-decision itself, an area around which many observers experience anxiety.

### **6.2.3 What elements of 'effective teaching' do observers focus on in evidencing their OTL decisions?**

A useful and significant aspect of this research has been the opportunity to capture the range of judgement-criteria cited by observers in a large number of observations: real-life judgements derived from real-life educational contexts. Analysis of this data facilitated: the compilation of a list of frequently occurring judgement-criteria; tables demonstrating the clear relationship between the frequency of these criteria and the grades awarded; and the creation of separate profiles of criteria that appeared to relate to different levels of OTL grade, achievement and attendance rates.

From the findings, it was possible to derive a list of the elements of ‘effective teaching’ most privileged by observers, as either strengths or improvements:

- Differentiation
- Feedback
- Planning
- ICT
- Questioning Technique
- Pace
- Teaching Methods

At first glance, this was not, perhaps, a surprising list. It appeared to offer a ‘common-sense’ collection of basic ‘teacherly’ virtues. However, a closer analysis reveals a more complex picture of the foci of, and influences on, the observer. ‘Planning’ and ‘Pace’ were, indeed, elements of ‘effective teaching’ as old as time and ‘Teaching Methods’ was a portmanteau category of different classroom approaches. However, the concentration on ‘Differentiation’, ‘Feedback’, ‘Questioning Technique’ and ‘ICT’ required closer attention. These four criteria demonstrated a tendency to reflect prevailing educational theory, and ‘fashion’.

‘Feedback’ and ‘Questioning Technique’ were two of the key elements of Assessment for Learning (AfL) introduced to mainstream teaching by Black and Wiliam (1998), becoming a standard approach in schools during the 2000s. These ideas took longer to reach the ACE sector and have, in the last ten years, been the subject of considerable in-service training at Newbold.

‘Differentiation’ has had a longer history, being mentioned by Shulman (1987), for example. The prevalence of this criterion - it is the single most commonly cited - might

be explained by the centrality of mixed-ability teaching within Newbold. Planning of differentiated learning activities and the production of materials targeting the individual needs of learners was a very central concern; it was practically one of the ‘values’ of the service. The presence, or lack, of proficiency in ‘Differentiation’ was, therefore, unsurprisingly highly privileged in observer decision-making.

The high placing of ‘ICT’ might have appeared to reflect the importance of e-learning in modern approaches to teaching. However, it is worth noting that, for a time, the use of ICT was an institutional priority and had been the subject of considerable emphasis, with examples of the use of ICT being collated from observation reports for official use. It was, therefore, natural that it became routine to note the presence, or absence, of this element during observations – despite widespread feeling that it was not a real indication of effective teaching. This was a prime example of institutional priority influencing the grade judgement.

Evidence from the data also confirmed findings in the literature (Sadler, 1985) suggesting that observers operated a ‘hierarchy of criteria’, in which some aspects were valued more highly than others: ‘Differentiation’ being considerably more important to observers than ‘Paperwork’, for example.

### **Criteria and grades**

The findings demonstrated a clear relationship between OTL grade-decisions and the occurrence of frequently cited judgement-criteria. **Table 5.3** (p144) showed that the frequency of particular strengths, for example ‘Differentiation’, could be seen to increase with grade and areas for improvement appeared to diminish. There is also a consistency within the findings as to the strengths and improvements applied by observers to their grade judgements within the levels. This suggests that analysis of the

judgement-criteria would be more important and useful, in terms of addressing QA and facilitating CPD, than the raw grades themselves.

Significantly, in my opinion, this led to the discovery that specific profiles of judgement-criteria could be found for each of the observation grades (see section 5.4.5, p144) as well as levels of accreditation and learner attendance (section 5.5.4, p153). These profiles of criteria appeared to highlight the differing aspects of classroom teaching associated, in the observers' judgement, with different levels and outcomes of 'effective teaching'.

#### **6.2.4 How do observers' judgements relate to other potential outcome measures of 'effective teaching'?**

As discussed in **Chapter 2** (see p57), recent research has questioned the efficacy of OTL as an assessment of 'effective teaching', in terms of the relationship between grade judgement and learner achievement (Coe et al., 2014; Strong et al., 2011). That is, when learner achievement has been used in studies as the 'objective' measure of 'effective teaching', then OTL appeared unable to differentiate the 'more effective' teachers from the 'less effective'. I was interested in investigating any relationship between OTL and other potential outcome measurements of 'effective teaching', principally achievement and accreditation, but also rates of learner attendance.

##### **Accreditation rates**

In agreement with the recent literature mentioned above, the findings clearly indicated that there was little relationship between the OTL grading and overall rates of learner achievement. However, a clear difference was shown at the boundary between the

‘acceptable’ grades (Grades 1 and 2) and the ‘unsatisfactory’ (Grades 3 and 4). This appeared to confirm OTL’s usefulness in terms of determining basic competency; grading did relate to achievement, at least to this extent.

However, this finding offered little support for the operation of four distinct levels of ability, and might appear to support the current moves towards abandoning the grading of OTL (Ofsted, 2015). The fact that accreditation rates did not indicate a difference between OTL Grades 1 and 2, might suggest that, when making judgements about teaching quality, observers were focussing on qualities quite different to those linked to achievement of accreditation, with qualitative differences between the top two grades being based on other factors relating to the observers’ conception of ‘effective teaching’ - and that conception was not identical with high learner achievement alone.

### **Attendance rate**

It was entirely possible that there were all sorts of alternative factors that played a part in influencing attendance in the classes under consideration here. However, given that some tutors were more consistently able to secure higher rates of learner attendance than others, it seemed reasonable to conclude that teacher-factors are, at least, partly responsible. The consensus within the small collection of studies within the literature to deal with the question (Davidovitch and Soen, 2006; White, 1992; Garner, 2006) supported a relationship between teacher effectiveness and attendance rates.

Perhaps surprisingly then, the findings indicated that there was no relationship between OTL grades and attendance rates. This finding might suggest that it would be unwise to include attendance as a factor in any post-grading assessments of ‘effective teaching’.

### **6.3 Profiles of judgement-criteria**

An interesting finding was that different conceptions of ‘effective teaching’ – excellent classroom performance; high learner achievement; high attendance rates – appeared to relate to different profiles of teacher skills, abilities and attributes – as reflected in the judgement-criteria in the observation reports.

Whilst there was no clear relationship between OTL grades and either the overall achievement rates, or the rates of attendance, there was, however, an emergent descriptive relationship between these outcomes and specific profiles of judgement-criteria: the suggestion that a particular set of strengths, for example, might link to a practitioner’s ability to help learners to achieve more highly, or come to class more consistently.

These different profiles of judgement-criteria appeared to indicate that different sets of strengths and improvements might be responsible for different outcomes in the classroom: for example, ability with ‘Planning’ and ‘Resources’ being more important in achieving accreditation; skills of ‘Differentiation’ and ‘Classroom Management’ more responsible for high learner attendance. It was also interesting to note that the profile associated with achieving the highest OTL grade, supposedly the most ‘effective’ teaching, was more similar to the high attendance profile, than it was to that of the highest accreditation. This may suggest that those qualities most appreciated by learners, which keep them attending more regularly, were similar to those most highly prized by observers.



## 6.4 Consistency between findings

Material for consideration in this study originated from three different, although not unrelated, areas: primary data sources (interviews, questionnaires), the analysis of secondary data (observation reports, achievement and attendance records), and more subjective ethnographic reflections. A concern may arise as to the degree to which there was consistency between these elements.

This section will consider the issue of consistency: first as it occurs between the primary and secondary research findings, and then between the research overall and the ethnographic reflections in **Chapter 2**. I will also comment on the methods used and sources accessed in relation to this issue.

### Primary vs Secondary research

There was a clear divide in terms of the nature of the primary and secondary research in this study. The primary data was collected via interviews and questionnaires concerned with observers' attitudes towards and beliefs about OTL. This data was based on what the observers in each sample said, remembered or thought. It was subjective in nature and based on opinion and belief, as well as experience. The secondary sources provided data concerned with the outcomes of actual observations, and for comparative purposes there was also data on learner attendance in classes and learner achievement. There was a clear connection between these two phases of my research: the interviewees and respondents to the questionnaires were a subset of the observers who carried out the observations and compiled the observation reports that formed the key element of the secondary data source. This fact in itself might be thought to ensure some consistency.

It was found that many of the elements of 'effective teaching' mentioned by the subjects of the primary research, also appeared in the observation reports (as well as

closely reflecting the list of elements found in the literature, see p53). There was little disagreement about the most important: 'differentiation', 'planning', 'questioning technique', and so on. There was a telling belief amongst the interviewees that 'learner voice' represented an important factor in their decision-making (see p139), but that is not reflected in its rate of inclusion as a strength or improvement in the reports.

However, this may indicate that picking up on what learners have to say about their teachers plays its part informally in helping observers to form their judgements, even if it is rarely included as a criterion.

Other criteria of judgement, as revealed in the interviews, also failed to feature in the secondary data. For obvious reasons, the observers' revelation that personal relationships between themselves and their observees can be a factor in their judgement (see section 5.3.3 p131) does not appear on the observation reports. This is a function of the more intimate nature of the interview, and to a lesser extent the questionnaire, whereas the observation report is an official document and subject to scrutiny by senior management. The primary data can reveal elements of the judgement process that the secondary data cannot.

The analysis of accreditation (see section 5.5.2, p150) and attendance rates (see section 5.5.3, p151) allowed a comparison to be made between these 'outcome measures' of 'effective teaching' and the grades (and judgement-criteria) attributed to the same teachers in their observations. There was considerable inconsistency here, if it was expected that 'effective teaching' would lead to high levels of achievement or attendance in a direct and unproblematic way. However, the interview data clearly indicated that the view that observers themselves took of 'effective teaching' was far more complex and contingent on a wide range of external and personal factors. In this respect the

findings of the primary data offers an explanation of the apparent contradiction in the secondary data.

### **Research vs Ethnographic Reflections**

My own beliefs and feelings about OTL were informed by my previous, and on-going, experiences of observing and being observed, as recorded in the ethnographic reflections in **Chapter 4** (p92), and it would be reasonable to assume that, although highly personal, these experiences may not be altogether unique – they were, indeed, reflected in the literature on the subject (O’Leary, 2013; Boocock, 2013; Cockburn, 2005). It might also be assumed that such experiences could affect an observer’s understanding of, and approach to, the observation of others: inclining such an observer, perhaps, to a more empathetic awareness of both what their observee might be going through, and how they might decide to try and present their practice in the best light in order to receive the affirmation of a higher grade. These assumptions are supported by the findings of the interviews in which most observers expressed their concerns for their observees’ feelings and the consequences their judgements might have – in some cases these relational factors altered the grade awarded to the observee (see section 5.3.3, p131).

From my reflection on the experience of observing a tutor (see section 4.2.2, p105), it became clear that the act of coming to a judgement was a near-automatic, instinctual phenomenon. This was in line with theories of expert decision-making in the literature (Hastie and Dawes’ (2010, Brown and McIntyre, 1993) and was consistent with the findings of the primary data that showed that all the observers asked came to initial judgements as some kind of a “gut feeling” (see section 5.3.1, p127).

My experiences of co-observing were illuminating in that they revealed aspects of the decision-making process that are usually private to the individual observer. These revelations were remarkably consistent with the findings of the primary research. For example, my experienced colleague (see section 4.3.1, p111) showed an empathetic focus on the well-being of her observee, a concern that the experience should not be overly stressful or humiliatingly negative, these concerns were also a feature of the data arising from the interviews (see section 5.3.3, p131) and help to illuminate the more ‘authentic’ path to judgement some observers take which attempts to account for their knowledge over time of the observee’s ability and previous performance, and which some believed constituted a more accurate picture than a one-off ‘snap-shot’.

The inexperienced observer’s struggle with her grade-decision (see section 4.3.2, p113) illustrated the contingent nature of the whole judgement process: although sure of her own opinion, based on her experience as a practitioner, she was prepared to compromise her grade-decision on the basis of factors external to the observation itself. This suggested that the relationship between grade-decisions and ‘effective teaching’ might not be as direct as otherwise expected. The contingent and complex nature of the grade decision revealed here was consistent with the findings based on the secondary data, specifically the lack of any clear relationship between the different potential measures of ‘effective teaching’.

It appears clear that the ethnographic reflections in **Chapter 4** offer more than mere background information and context. The nature of the insights arising from these reflections appears generally consistent with the research findings, especially from the primary sources which are themselves qualitative in nature. The primary research methods – interviews and questionnaires – may have been more rigorously applied and controlled in accordance to accepted methodology as opposed to the looser and more

‘subjective’ ethnographic material, but the nature of the findings in each is suggestive of a consistent picture of the lived experience of OTL.

## **6.5 The relationship between OTL decision-making and the concept of Wisdom of Practice**

### **6.5.1 Introduction**

In all the examples of OTL decision-making illustrated in the ethnographic reflections, the nature of the judgement itself was seen to be intuitive in the first instance: my own decision came almost automatically; the experienced observer’s decision just ‘came’ to her; the newly-appointed observer’s was ‘gut instinct’ based on years of experience in the classroom. These findings were all suggestive of the characteristics of the nature of Wisdom of Practice *in situ*, discussed on page 47.

My experience of closely reflecting on my own OTL decision-making (see section 4.2.2, p105) revealed that my judgement was fixed very quickly, only minutes into the observation, and that it was based on a ‘feeling’ rather than any kind of coherent conscious weighing-up of factors. This was strongly suggestive of the kind of experience-based, internalised phenomenon that has been identified in the literature as Wisdom of Practice.

Although ultimately an abstract phenomenon, and difficult to subject to evidential analysis, the concept of Wisdom of Practice may nevertheless be usefully applied in the light of these results. It is, at least, a convenient way of referring to the amalgamation of theory, experience and background knowledge that observers bring to bear on their decision-making. The term did appear to effectively describe an observable facet of

professional life, as seen in a wide range of studies in the literature (Shulman, 2004; Higgs, 2012; Lunenberg and Korthagen, 2009).

To adopt Leinhardt's (1990) definition, the Wisdom of Practice can be thought of as:

“the wealth of teaching information that very skilled practitioners have about their own practice ... deep, sensitive, location-specific knowledge of teaching”  
(Leinhardt, 1990, p18)

Access to the Wisdom of Practice, therefore, will be behind a great deal of what the experienced practitioner does, despite not necessarily being able to consciously address it (Yeager, 2000). All of the observers, here, were highly experienced teachers as well; and so Wisdom of Practice might reasonably be presumed to be available, and useful, to them in making OTL judgements.

### **6.5.2 Evidence for Wisdom of Practice**

The findings appeared to support the contention that Wisdom of Practice was involved in OTL decision-making.

**Table 6.1**, below, compares characteristics of Wisdom of Practice taken from the literature (see section 2.6, p43) with examples, from the findings, of what observers said about their decision-making. It can be seen that OTL decision-making appeared to correspond to the basic characteristics of the Wisdom of Practice:

<b>Characteristic of Wisdom of Practice</b>	<b>Example Evidence from findings</b>
Automatic	<p>“It’s just there; you don’t have to think about it.” (Interview D)</p> <p>“Largely instinctive” (Interview A)</p>
Intuitive	<p>“I have some sort of gut feel for the grade I think it deserves.” (Questionnaire 8)</p> <p>“Basically a best guess” (Questionnaire 4)</p>
Experiential	<p>“Weighted with the experience over many years of observations” (Questionnaire 4)</p> <p>“I have an idea in my mind of what a ... good lesson is” (Interview C)</p>
Hard to explain	<p>“Obviously the feeling comes from somewhere...” (Questionnaire 7)</p> <p>“Who is to say I am right?” (Interview B)</p>

*Table 6.1: Characteristics of the Wisdom of Practice compared to findings*

As discussed above, in line with studies of expert knowledge in the literature (Dreyfus and Dreyfus, 1980; Brown and McIntyre, 1993; Ainley and Luntley, 2007; Hastie and Dawes, 2010), observers often found it difficult to explain their decision-making process, and much of the time their initial judgements were described in terms of holistic automatic thinking. This finding might imply that observers are accessing something like what has been termed the Wisdom of Practice; it is clear that observers, themselves, believed that it was in their ‘experience’ that the principle source of their

decision-making is located. That they were often troubled by doubts about their decisions might offer supporting evidence of observers' lack of conscious awareness of their decision-making: they arrived at a decision, but they did not fully understand how or why; doing so based on a deep-seated automatic access to their experience-based Wisdom of Practice. To return to Hastie and Dawes' (2010) analogy, observers reached their initial OTL judgement in much the same way that an experienced driver, whilst deep in thought or conversation, might arrive safely at an intended destination with little memory of actually having driven there.

The inclusion of the concept of Wisdom of Practice might be further justified on the basis of the wide degree of agreement as to what constitutes 'effective teaching' found amongst observers and a wide range of sources in the literature. **Table 6.2**, below, summarises and compares the most frequently occurring judgement-criteria from my data with elements of 'effective teaching' most commonly identified in the literature. The comparison clearly demonstrates a high degree of agreement in terms of inclusion: 13 of the 20 most frequently occurring judgement-criteria in the observation reports are also among the 20 most common elements found in the literature (blue shading). Overall, 45 out of the 56 judgement-criteria are also found in the literature.



Ranking in observation reports	Judgement-criteria in observation reports	Ranking in literature
1	Differentiation	11
2	Feedback	3
3	Planning	2
4	ICT	N/A
5	Activity	5
6	Questioning technique	14
7	Classroom management	13
8	Pace	N/A
9	Resources	N/A
10	Teaching methods	16
11	Subject content	10
12	Aims	6
13	Peer working	N/A
14	Reflective practice	15
15	Paperwork	N/A
16	Accreditation focus	N/A
17	Clear instructions	8
18	E&D	N/A
19	Assessment	1
20	Engagement	9

*Table 6.2: Comparing the ranking of most frequent 20 judgement-criteria in observation reports with frequency of occurrence in sample of sources from literature review.*

Judgement-criteria from the observations not found in the literature (red shading), indicate such local and contextual variations as management priorities, paradigmatic shifts in inspection foci, and other external factors. ‘ICT’ and ‘E&D’ were amongst more recent priorities excluded in older studies; ‘Accreditation Focus’ and ‘Paperwork’ might be considered, as they are by many practitioners, external to the domain of ‘effective teaching’ – and included in observation reports due to their status as management/institutional priorities. An individual’s, or a particular sub-group’s, Wisdom of Practice appeared able to evolve and respond to localised contexts, whilst maintaining a strong core of values that appeared relatively universal within the field.

It was the almost instinctual nature of OTL decision-making, based on this shared bank of ideas, views and beliefs about what comprises ‘effective teaching’, that laid the foundation for the claim that OTL may constitute an expression of the Wisdom of Practice in action. Observers accessed this internalised set of values in order to form their initial judgements, before reaching more consciously for evidence based on judgement-criteria drawn, to an extent, from the same shared source.

## **6.6 “How do observers make judgements in the Observation of Teaching and Learning?” - A model**

My overarching aim, in this study, was to address the question of how observers make their OTL judgements. To this end, I have posed my four research questions and reported on the findings of my research. It was my task to attempt to present as clear an explanation as possible of what I believe happens when observers come to their decisions. I have chosen to do this in the form of a theoretical model (see **Fig. 6.4**, p185): theoretical in the sense that it attempts to be a “‘way of making sense’ of a certain slice of the empirical world”, in this case OTL, and will hopefully “help one understand some social process” (Abend, 2008, p178): the way in which observers come to their OTL decisions.

This model is a schematic simplification of a human process that must be infinitely more complex – and messy – than it is made to appear. Clarke and Primo (2007, p751) argue that models are like maps: “representations of reality” which are “partial”, in that they “represent some features of the world and not others”. In response to this partial representation, the key question to ask of a model is not whether it is true or accurate, but rather whether it can be appropriately “used for specific purposes” (ibid).

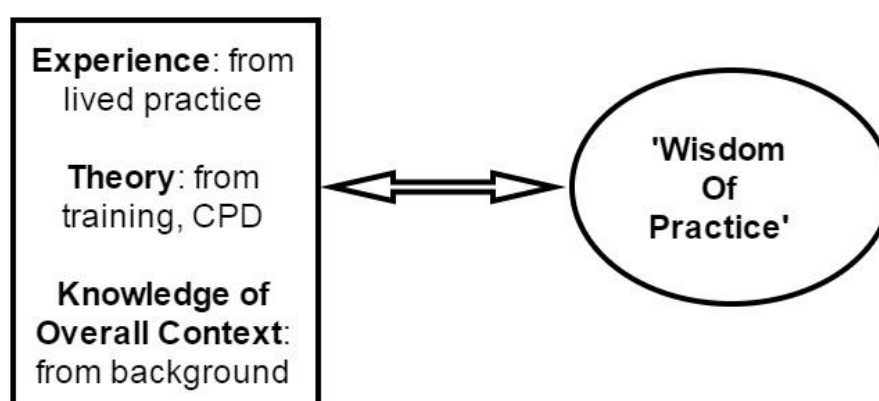
The specific purpose of the model below is not an attempt to map exactly what happens when an observer considers an OTL decision. It might, however, provide a useful lens through which to focus on some key elements of that decision-making process, and the probable relationships between those elements.

This model might be the first that attempts to explain the way observers reach OTL decisions. For this reason, it remains to be seen if it will prove useful in facilitating further research and understanding in this field. It is an attempt to draw attention to the multifaceted nature of the OTL judgement process which is revealed in the findings. It acknowledges the dual approach to decision-making suggested in the interviews: the holistic instinctive decision and the heuristic criterion-based judgement.

The model that follows was initially suggested by the work of Lunenberg and Korthagen (2009), discussed in **Chapter 2** (see p44-45), and in particular, their ‘triangular relationship’ between Practical Wisdom, Theory and Experience (p229, fig.1, reproduced on p45). Whereas this ‘relationship’ placed, at its centre, the teacher (and all that she experiences and does in the classroom), in my model, I have sought to locate the OTL judgement in relation to the observer’s Wisdom of Practice. Lunenberg and Korthagen’s triangle then becomes a facet, or starting point, of the whole judgement process – a way of incorporating the Wisdom of Practice into the process. In order to do this, I adapted the original diagram as an interim stage towards constructing the model itself.

In adapting Lunenberg and Korthagen’s ‘triangular relationship’ (see **Fig. 6.3** below), ‘Theory’ and ‘Experience’ have been preserved as factors in relation to the what I prefer to call the ‘Wisdom of Practice’ (following Shulman (1983), a more usefully inclusive concept, in this context, into which the various facets contribute, rather than co-existing in equilibrium). However, this new model subdivides the category of ‘Experience’, by

introducing a fourth factor: ‘Knowledge of Overall Context’. I have found it useful to separate ‘Knowledge of Overall Context’ from ‘Experience’ in order to highlight the importance of contextual understanding in informing judgements based on specific aspects of practice: for example, the observers’ knowledge of their own institutional processes. This subdivision incorporates the distinction in Shulman’s (1987) ‘Categories of the Knowledge Base’ (see section 2.6, p44): specifically, ‘knowledge of learners and their characteristics’ and ‘knowledge of educational contexts’ which includes an understanding of particular groups of learners in particular institutions. This was reflected in the findings: although observers strongly located their decision-making in their accrual of ‘experience’, when they talked about ‘experience’ they appeared to distinguish between their generalised collection of applied skills, know-how and understanding and their knowledge of particular settings and situations.



*Fig.6.3: The relationship between experience, theory, knowledge of overall context and 'Wisdom of Practice' (adapted from Lunenberg and Korthagen, 2009, p229, fig.1)*

Another major difference here is that I have rendered Lunenberg and Korthagen’s ‘triangular relationship’ as more linear. I believe this is appropriate, as I was not looking to express a definitional outline of the Wisdom of Practice itself, or its place in the classroom experience, as were Lunenberg and Korthagen; rather I was focussed on the way these factors feed-in to the process of OTL decision-making. I was, therefore,

more interested in the forward flow from the observer's frame of reference (including Theory, Experience and Knowledge of Overall Context) through Wisdom of Practice towards the final OTL judgement. I have, however, maintained the original two-way relationship between these aspects, as I acknowledge the essentially unstable and evolving nature of Wisdom of Practice, with new experiences constantly feeding into and back from each other to create, adapt and reinforce understanding.

### 6.6.1 The model

The real-life decision-making process of an observer is a complex coming together of various internal and external factors. This process will, inevitably, be more personal, fragmented and complex than any single model could possibly describe. In fact it would be detrimental to the purpose of a model to even try to include all the various permutations of possible factors that might come to bear on a single decision. What I have attempted here is to propose a stylised outline of a much more complex process. My intention was to indicate certain facets of the process that have arisen from the findings and suggest how these factors might influence the final OTL decision.

There follows a description and explanation of the various facets of the model:

**Overall:** The elliptical space at the centre of the diagram represents the OTL Decision-making Process as a specific instance of classroom observation; those elements appearing within this space should be considered as internal to the observer's experience of considering the observed teaching session. The elements occurring outside the elliptical space should be considered external to the specific instance of observation; these are elements that are brought to bear on the observer's judgement process, and they originate from before the specific instance of observation.

(1) **Observer's Frame of Reference:** these factors, external to the direct experience of the classroom observation, arise from the observer's lived **Experience**, both of classroom teaching and of carrying out previous observations (see section 5.3.4, p133); her accrual of **Theory** from various sources including from Ofsted (see section 5.2.4, p123), ongoing CPD and interaction with other practitioners (see section 5.3.2, p131); and her wider **Knowledge of Overall Context** in which the observed teaching is taking place: for example, the structure and culture of the institution and the sector; the make-up and background of the student body, and so on (see section 5.2.6, p125).

As discussed above, these factors exist in a relationship of mutual feedback with the observer's access to the 'Wisdom of Practice' (2). The three strands of the Frame of Reference inform and define the Wisdom of Practice; however, as represented by the two-way arrow between them in the model, each experience of OTL decision-making will also add to the store of experiences that constitute the Wisdom of Practice and will inform all the elements of the Frame of Reference

(2) **'Wisdom of Practice':** This new model locates the Wisdom of Practice at the heart of the Classroom Observation; it is the key factor in the decision-making process, although neither the sole factor, nor one uninfluenced by other factors. In this model, the Wisdom of Practice is a construct related to and arising from the individual observer's Frame of Reference (1); and it is this construct that will be the principle location of the Initial OTL Judgement (3). This was the judgement that observers experience as the holistic 'gut feeling' instinctual decision (see section 5.3.1, p127).

(3) **Initial OTL Judgement:** As seen in the findings, observers' initial decision-making was holistic and intuitive (see section 5.3.1, p127); their judgements based on an

internalised standard which they found difficult to rationalise explicitly (see quotes, p128). In this model, such intuitive and internalised decision-making is presumed to arise from the Wisdom of Practice.

(4) **Verification and Referencing:** In order to verify, or perhaps ‘fine-tune’, the initial judgement – an essential part of the process considering the widely expressed doubts observers experience in relation to their decision-making – other externally referenced factors are then brought to bear, more consciously and deliberately than the Wisdom of Practice. In this model, I propose that two simultaneously effective groups of factors are considered in order to test or revise the initial judgement:

(4a) **Judgement-criteria:** observers stated that they considered their personal ‘check-lists’ of judgement-criteria in order to check the judgements they already had in mind (see section 5.3.2, p129). This list of criteria (see section 5.4.2, p139) – which will eventually feature as strengths and improvements on the observation reports – originated from numerous potential sources. In this model, one prominent source is presumed to be the observer’s own Wisdom of Practice, as criteria arise initially from their experiential evaluations: ideas already held about what constitutes ‘effective teaching’ (Sadler,1985). Other sources include: Ofsted inspection criteria; new trends in theory emergent from training; and the exigencies of institutional targets such as preparation for accreditation (see section 5.2.6, p125).

(4b) **Situational Factors:** these are both the long-term environmental and contextual realities within which the observed teaching takes place, as well as more temporary or emergent issues that might have affected the particular incidence of observation. Examples might include: a failure of ICT; the unexpected absence of a classroom assistant; an over-crowded room; or on-going relational issues among members of the class (see section 5.3.5, p133). A key factor identified by observers was the question of

their relationship with their observees: this is acknowledged to have a potential effect on the judgement process (see section 5.3.3, p131). The coming together of the judgement-criteria and the situational factors will be mediated through and informed by the Wisdom of Practice – and the subsequent revision or reinforcement of the initial OTL judgement will still, therefore, be subject to the influence of the observer's Wisdom of Practice.

(5) **Interim OTL Judgement/s:** having checked and fine-tuned their Initial OTL Judgement (3) via the Verification and Referencing phase (4), observers will now arrive at an interim judgement which, although still significantly based on the holistic decision emergent from their Wisdom of Practice, will now be a more nuanced construction connected to perceived situational and shared external factors.

It is likely, given the levels of self-doubt expressed by many of the observers, that the OTL decision will pass through a number of cycles of re-consideration and interim judgements (see section 5.3.6, p134).

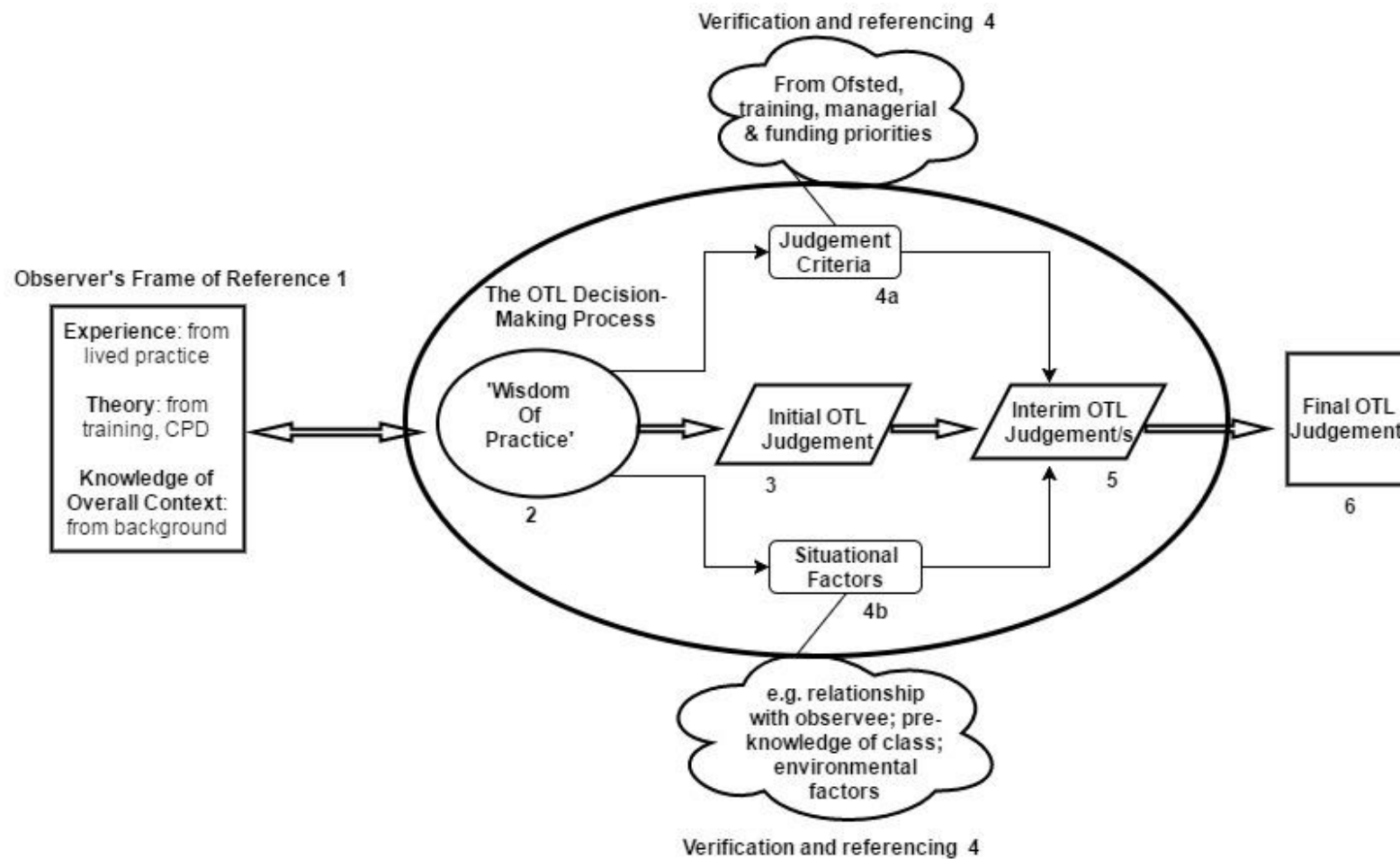
(6) **Final OTL Judgement:** The observer delivers the grade. This grade, and the set of judgement-criteria, categorised as strengths and improvements, concluded to be consistent with and evidential of it, constitutes the outcome of the OTL process. It is an important outcome: an official statistic with high-stakes consequences for the observee, and significant implications for the institution.

It should be pointed out that the grade is a required end-point of the OTL process described in this model. The observer is obliged to produce the grade and the process cannot fail in its delivery. The judgement process described by the model will continue to its outcome regardless of how compromised the process becomes. The fact of the grade decision then self-justifies the process – it is evidence that the observation has



delivered its satisfactory conclusion. The observer may have had doubts about her decision (see section 5.3.6, p134), may even question the value of the grading process itself (see section 5.2.5, p124), but the grade will stand.

At the end of this complex and opaque process, a contingent and necessarily imperfect attempt to bring together the various strands of judgement has resulted in a public expression of one experienced practitioner's opinion about the 'effectiveness' of another practitioner's teaching. That expression, the grade, will then be used by others – management, Ofsted – as 'evidence' of quality assurance without any reference to the specific personal, professional and contextual factors that went into producing it.



**Fig. 6.4:** "How do observers make judgements in the Observation of Teaching and Learning?" A model (partly suggested by Lunenberg and Korthagen (2009, p229, fig.1))

### **6.6.2 Critique of the model**

Although I was confident that this model emerges logically and convincingly as a consequence of the findings, I have identified three potential criticisms:

Over-simplification is potentially an issue, and one that might be addressed following further refinement and testing. However, I believe the level of clarity in the current model is valuable in terms of explanatory usefulness.

Secondly, I cannot pretend to offer a psychological or neurological explanation of what goes on within observers' minds, as I have not trained within those traditions. The process illustrated derives from observable elements within the behaviour of the observers, and are based on the findings of the interviews and the questionnaires – the only really abstract element is the Wisdom of Practice (as discussed above).

Finally, the model does not capture the dynamic nature of a real-time observation event. However, this is no more true than any map can authentically feel like a journey in the real world – but maps have their purposes and can, like this model, be illuminating and instructive.

## **Chapter 7 – Conclusion**

### **7.1 Introduction**

‘How are judgements made in the Observation of Teaching and Learning?’ At first glance, providing an answer to such a question might appear a modest ambition. However, this impression would be deceptive. I have found that even those practitioners whose job it was to make these judgements, struggled to explain how they make them, hold contradictory views, and harbour doubts and anxieties about the decision-making process. Indeed, as an observer-manager myself, with years of experience of making OTL judgements, a major factor in my decision to approach this area of research was that I was unsure how it actually worked. I believe that the findings of this research represent a step towards understanding this far from transparent process.

In this chapter, I present a summary of the main findings, discuss what I consider the strengths (and possible limitations) of my research, and make claims to what I believe to be its contribution to knowledge. Further, I comment on the implications of my findings and offer recommendations for practice and possible further research.

### **7.2 Overview of the research**

As stated in **Chapter 1**, my purpose was to shed light on the decision-making process at the heart of OTL by addressing the following key research questions:

- I. What do observers believe about the OTL process?
- II. How do observers explain their OTL decision-making?

- III. What elements of ‘effective teaching’ do observers focus on in evidencing their OTL decisions?
- IV. How do observers’ judgements relate to other potential outcome measures of ‘effective teaching’?

In addressing these research questions, I have:

- presented qualitative findings based on interviews and questionnaires addressing what observers said about the processes behind their OTL judgements
- used the findings of analysis of secondary data to explore the relationship between OTL grade-decisions and the judgement-criteria used as evidence for them
- attempted to operationalise personal experiences of the OTL process in order to provide background, context and an insider’s perspective
- attempted to form theory regarding the role of the concept of Wisdom of Practice in the OTL process
- presented a proposed theoretical model of the OTL judgement process

### **7.3 Summary of key findings**

#### **RQ1: What do observers believe about the OTL process?**

- I. Despite all reservations, observers believed that OTL was useful and important

- II. Observers recognised that OTL had both a strong QA purpose as well as a strong CPD purpose; and that these two main purposes were often in tension
- III. All observers were critical of the grading process and had concerns about the intrusion of institutional targets and managerial priorities on the usefulness of OTL

**RQ2: How do observers explain their OTL decision-making?**

- I. Although apparently contradictory, observers appeared to use both intuitive/holistic and criteria-based decision-making methods
- II. Observers admitted that their OTL decisions were affected by their relationships with observees
- III. Observers cited their own professional experience as a basis for decision-making
- IV. All observers expressed doubts about their OTL decisions

**RQ3: What elements of ‘effective teaching’ do observers focus on in evidencing their OTL decisions?**

- I. Judgement-criteria were offered as the main evidence for OTL decisions.
- II. Secondary data analysis made it possible to determine the most frequently occurring judgement-criteria cited by observers at each observation grade level – with each grade associated with a different profile of specific criteria.
- III. There was a clear relationship between OTL grade and the incidence of frequently cited judgement-criteria as strengths and improvements.

**RQ4: How do observers' judgements relate to other potential outcome measures of 'effective teaching'?**

- I. Achievement and accreditation rates had no observable relationship with OTL grades overall - but a difference was observed at the boundary of the acceptable OTL grades
- II. Attendance rates had no observable relationship with OTL grades
- III. Other outcome measures were associated with different profiles of specific judgement-criteria

## **7.4 Strengths**

In this section, I briefly discuss the elements of the thesis I consider to be particular strengths, as well as those that might offer a contribution to knowledge.

Overall, one of the most significant strengths is one of context: OTL is a relatively under-researched area; and it is certainly the case that very little consideration has been previously made of the decision-making process behind classroom observation.

### **7.4.1 Focus on the observer's OTL experience**

As explained in the review, there was little in the literature directly relating to the observer's experience of OTL. The intention of this study was to address this quite specific gap. The study found that observers experience doubt about their decision-making; that observers appear to access both holistic and heuristic decision-making; that their own experience is cited as the major factor in their decision-making.

## **The impact of the observer's relationship with the observee**

Another important finding was that many observers allowed their OTL decision-making to be affected by their relationships with their observees – the fact that they have known them across time and are generally aware of their skills and abilities – quite apart from the ‘snapshot’ nature of a single incidence of OTL. The importance of the relationship between observer and observee does not feature prominently in the literature, and this is one way the current study makes a contribution to knowledge.

Some in-house observers felt having access to knowledge and understanding of their observees’ teaching was highly positive, in that this longer term knowledge of their observees allowed a more ‘authentic’ appraisal of their ability. Others worried that this might lead to bias, but most accepted the effect of the relationships as generally positive. It can be argued that this relational knowledge forms a key part of the observee’s “professional capital” (Hargreaves and Fullan, 2012) and that it would be understandable and defensible that they should make use of it in arriving at their OTL judgement.

In theory, observers were constrained to produce a judgment based on a single example of observed teaching, and the observation report and final grade was put forward officially on this basis. However, it was both inevitable, and arguably desirable, that previous knowledge of the observee would be used to bring wider focus to the decision. That appeared to be what happened in many cases according to the findings of this research.

In managing this tension between what can be observed and what was actually known, the observers appear to exercise “principled infidelity” (Wallace and Hoyle, 2007, p19). In other words, they adapt “the imperatives of policy to the contingent needs” (p18) in



order to arrive at a ‘fair’ and effective OTL judgement. In Wallace and Hoyle’s terms, the ‘infidelity’ rests on the fact that the observer does not “slavishly adhere to expectations”(p19) – that is the official way of carrying out the observation on the ‘snapshot’ model – but the approach is ‘principled’ because they “seek to sustain their professional values” (ibid).

#### 7.4.2 Ethnography

I believe that the ethnographic and auto-ethnographic work, detailed in **Chapter 4**, constituted some of the most important elements. The value of this work lies in the fact that it collects together, and reflects on, authentic examples of the type of subjective lived-experience that observers engage with during the OTL process.

Although subjective and potentially anecdotal in nature, these reflections offered important insights; especially illuminating background and context for readers new to the subject.

The key ‘findings’ of the ethnographic section were as follows:

- The automatic and instinctual nature of the initial OTL judgement. This supports the contention that Wisdom of Practice is at the heart of OTL decision-making.
- Factors outside the boundaries of the observed lesson are important to the final grade judgement: ‘Situational Factors’
- The relationship between the observer and the observee can directly affect the grade judgement.
- The process of coming to a grade-decision is characterised by doubt.

- Judgement-criteria are secondary to the initial grade judgement. They are used to evidence and rationalise the grade-decision.
- Observees can influence the grade-decision through strategic preparation and presentation.

As one of my key claims is that Wisdom of Practice – an instinctual, automatic, largely unconscious access to knowledge and experience – is at the heart of OTL, it seems appropriate that an attempt has been made to capture various aspects of the OTL experience in a manner that did not rely on the conscious thought processes and recall of practitioners.

On a personal level, the auto-ethnographic reflection was certainly significant to me. It helped to place my thoughts on OTL into an immediate and, at times, surprising set of contexts: for example, I had not previously noticed the near automatic decision-making that characterised my own initial OTL judgement; I was given an insight into how I looked for and used judgement-criteria; I was alerted to the high level of strategic planning I undertook in order to prepare for being observed. An exercise that I had planned in order to give readers some background and context, proved crucial to my own understanding of my material. This new understanding was used to alter the direction of the research – especially, the interest in Wisdom of Practice - and, ultimately, underscores my conclusions.

### **7.4.3 Analysis of secondary data**

A key strength arises from the privileged access I had to a range of secondary data: chiefly, 137 observation reports pertaining to 71 observees created by 21 observers

(including the 6 respondents to the interviews and the 8 respondents to the questionnaires). This represented a wealth of data pertaining to observers' real-life OTL decisions, a virtually "complete" data set that pre-existed the research and the questions put to the observers in the interviews and questionnaires.

The most important and interesting information to arise from my analysis of these reports was the collation of the judgement-criteria cited by the observers in evidence of their grade-decisions. I do not believe any such collation has been previously undertaken.

**Table 5.1** (p139) represents, therefore, a useful contribution to knowledge, presenting, as it does, a list of the most common judgement-criteria applied to observations in the field, arising entirely from the practice of the observers themselves. These findings are obviously open to further interpretation: as to the origins of each criterion, for example. It should also be remembered that these observers, in the Newbold setting, used no official check-list of criteria, or observation schedule.

The analysis of these findings allowed me to show which judgement-criteria were most associated with the various levels of OTL grade. Care must be taken in the interpretation of these relationships, not least because the criteria cited and the grade-decisions arise from the same source in each case – the judgement of the observer – and, as we have seen, observers may have chosen judgement-criteria directly to justify the grade-decisions they had already made. However, assuming that both the grade-decisions and the judgement-criteria represent sincere attempts to deliver fair and authentic assessment – and I have no reason to doubt that this is the case – then the findings here are an interesting and useful insight into the nature of what observers consider 'effective teaching'.

This collation of judgement-criteria, and the comparisons afforded by cross-referencing OTL results with accreditation and attendance data, also offered the opportunity to construct a range of profiles of judgement-criteria: for each OTL grade, for observees with high rates of achievement and for high rates of attendance. These profiles revealed intriguing and somewhat unexpected insights: the most counter-intuitive finding being that observees achieving the highest OTL grades have quite different profiles of strengths than do those with the highest rates of achievement.

#### **7.4.4 The role of Wisdom of Practice in the OTL process**

It was a key interest to consider whether observers' OTL decision-making offered an example of the concept of Wisdom of Practice in action. I have concluded that it appears likely that observers access a 'bank' of experiences, theories and practical knowledge when they make OTL judgements, and this I have chosen to identify with the concept of Wisdom of Practice as discussed in **Chapter 2**. To an extent, observers appear to access this Wisdom of Practice 'automatically' and, at least, partly unconsciously. It was a strength that I was able to compare observers' elements of 'effective teaching', as emergent as strengths in the observation reports, with a survey of elements found in the literature (see **Table 2.5**, p54): from the high degree of agreement among practitioners, professional bodies, and researchers, it can be concluded that the Wisdom of Practice appears to be a 'shared' phenomenon – although constantly evolving for each practitioner, and subject to localised factors.

This claim to the involvement of the concept of Wisdom of Practice in the OTL judgement process is potentially a contribution to knowledge in this field.

#### **7.4.5 Model of the OTL judgement process**

By incorporating the concept of Wisdom of Practice, and in accordance with the findings of the interviews and questionnaires, I have presented a theoretical model of the observer's judgement process. I did so, partly, to bring together the various elements in a way that might clarify what had become, in my mind at least, a more complex process than might have been initially expected.

The model attempts to account for the complexity of the OTL decision, as experienced by many observers. It describes a process of refinement and reiteration that eventually leads to a decision, albeit one that many observers will have found difficult. This is accounted for in the model in the multiple pathways arising from the Wisdom of Practice and leading to the interim judgements: the Wisdom of Practice is not only linked to the holistic initial judgement, but also flows through to, and influences, the observers' interpretations of the various situational factors they might consider, as well as both helping to form and determining the application of the judgement-criteria in refining and evidencing their judgements. In this way, the Wisdom of Practice can be seen as the central element in the OTL decision-making process in the middle of a web of inter-related judgement factors.

#### **7.5 Limitations**

Despite the fact that the combined number of subjects for the interviews and questionnaires (14) represented a substantial sample of the total number of observers available to me in the ES department at Newbold (24), perhaps the main limitation was that some of my findings were based on a relatively small number of interviews and questionnaires. This was both a consequence of the financial and time constraints

inherent in carrying out a part-time professional doctorate, and a deliberate and positive decision to focus on depth rather than on the quantity of responses. It would have, perhaps, been advantageous to interview more widely. However, it is worth noting that the largely unanimous responses received from this set of subjects might indicate that a larger number of interviews may simply have resulted in more of the same.

It would perhaps have been interesting to have had the opportunity to carry out similar research – both primary and secondary – at more than one Adult Education institution. This might have offered more variation of situation and context with which to test my findings; and begun to address the question of whether the attitudes and practices discovered were universal in nature or particular to the setting. I was, however, restricted, again by time and financial restraints, as well as by my work-role, to just the one institution during the course of the study.

Another potential limitation was that one of the key foci - the Wisdom of Practice – appeared, due to its internal and subjective nature, somewhat resistant to direct analysis. I was, therefore, obliged to focus on outward signs of the characteristics of the phenomenon, as taken from the literature. That these match with details emerging from the interviews is, largely, the basis of my claim that Wisdom of Practice has a place in thinking about OTL decision-making. The ethnographic reflections in **Chapter 4**, especially those arising from my experiences of carrying out an observation, also appear to support the operation of the Wisdom of Practice in OTL decision-making.

It was unfortunate that I did not have the time or space to include a consideration of the views of the observees in this research. I did conduct a survey of 21 practitioners comprising one complete curriculum team within the ES department, the purpose being to enable comparison between the views of observees and the managers who observed them. I found a considerable degree of agreement between observees and observers as

to the key elements of ‘effective teaching’, with one significant area of apparent divergence: observees cited a range of highly subjective qualities not mentioned in observation reports’, for example, ‘honesty’ and ‘commitment’. These largely unobservable (in a brief classroom visit) elements were obviously vital for the relationships between teachers and learners, but were absent from the lists of elements cited by observers, perhaps precisely because they were difficult to observe. Ultimately, I made the decision to exclude this material because, although interesting, the results were not sufficiently relevant to answering my key questions. I include the table of results from the observee survey in **Appendix 7** (p226) for interest.

## **7.6 Implications and recommendations**

### **7.6.1 Theoretical**

Based on my findings, I make the claim that the Wisdom of Practice of observers is at the heart of the OTL decision-making process. The implications of such a claim are potentially important in terms of the understanding of the nature of OTL: what it is, what it could be, what it should be used for. I do not believe that the highly subjective nature of the concept should be considered a problem or a weakness in relation to OTL, as anything approaching objectivity could never really be established in such a field. Rather, the opportunity is there to embrace the claim for authenticity offered by a process based on the wealth of experience and knowledge of its practitioners, especially as, at the time of writing, the sector appears to be moving from summative graded OTL towards a more formative model of observation as part of CPD. Although supported by the evidence here, this is an underdeveloped theory that would benefit from further discussion, reflection and research.

My theoretical model (see **Fig 6.4** p185), as has been discussed, offers a view of OTL as a decision-making process that is primarily grounded in the experience of observers (both of their own teaching, and of previous observations). The thinking behind the model is discussed in section 6.6 (p177). However, I feel it is important to state here that I do not pretend that the processes of a human mind can be reduced to such a simplistic and stylised form. Rather, what I have attempted with this model is to represent some of the key findings of this research in a visual form that might usefully capture some of elements of the decision-making process, in order that they can be considered in relation to each other.

I see the model as a first step towards a more developed conception of how expert practitioners reach their decisions during classroom observation. Further research might be able to refine the various elements of the model, though it would be important not to compromise its utility by making it over-complicated. One fruitful area for future research might involve presenting the model to observers and observees and recording their feedback: do they recognise the systems? Does it make sense to them? I take the opinion that the value of a model of this sort is mostly dependant on whether it recognisably reflects real-life and it would be instructive to find out.

### **7.6.2 Practical**

In terms of developments within the practice, these findings appear to support moves, current at time of writing, to end grading in OTL (see section 1.5 p16). Although, inevitably, an element of the summative QA function would remain, this would most explicitly pertain to the observation of probationers and to the fraction of observations in which basic competency becomes an issue. In most cases, the shift in emphasis away



from grading might allow a rebalancing at the heart of OTL in the direction of the more qualitative and supportive CPD function that, according to the findings, most practitioners favour.

In practice, the collation of criteria may prove useful in terms of providing foci for strategic planning within institutions, or feed-in to CPD for individual practitioners and teams allowing practitioners to focus on the key skills associated with ‘effective teaching’, for example; or the different profiles of elements associated with higher rates of accreditation or attendance, offering insight into different strategic approaches. More generally, the data could find useful application in the preparation and training of classroom observers.

It is important to note that the research did not demonstrate a clear relationship between OTL judgement and the other putative measures of ‘effective teaching’ looked at in this study. Accreditation rates did not show an overall relationship with observers’ grading, it appears clear that observers are looking at qualities in the classroom that go beyond the requirements of merely teaching to the test. However, there appears to be no connection between the ‘effective teaching’, as judged by the observers, and rates of learner attendance. Based on this finding it would be a mistake to link attendance with conclusions about teacher effectiveness.

### **7.6.3 The future for OTL?**

I entered this area of research firmly believing that I would find results showing that OTL was inadequate as an assessment of ‘effective teaching’; that grading would appear subjective at best and arbitrary at worst; that there would be little connection between the findings of OTL and the outcomes of other indicators of teacher effectiveness.

However, having spent five years reading, researching, considering and reflecting on the issue, I reach the end of this journey in an entirely unexpected place. I now believe, unequivocally, that the observation of teaching and learning is an important and valuable process - possibly the only authentic way of capturing and valuing 'effective teaching' in the classroom.

When one experienced professional, accessing their wealth of knowledge, know-how and experience - their Wisdom of Practice - encounters another in their own field, then there is the truest opportunity of real meaningful assessment of ability. Some degree of effortful decision-making has always been involved despite the apparent flaws and shortcomings of the system.

As discussed in **Chapter 1** (see section 1.5, p16-17), the end may be nigh for graded OTL. Amid the clamour of the debate over the reliability of OTL, and calls to focus on other outcome measures of teacher effectiveness, such as accreditation-rates, there is a danger that OTL may become rejected as both inaccurate and undesirable. In this latter respect, its demise might well be assisted by practitioners, like my earlier self, who see only the overly judgemental and summative aspects; who regret the time-wasting and the diversion of focus from learner needs; who have experienced the raised anxiety that appears to be inherent in the practice. However, in my opinion, this would be a mistake, especially as the practice of OTL is in the process of transformation to a more developmental model.

This is not to say that I approve of grading, or can see any useful reason for it to continue, except perhaps in the case of competency procedures. However, the practice of grading has proved highly illustrative for the purposes of this research, allowing me to track the formation of observers' judgements in a way that a more formative process might not. For this reason, I was fortunate to have had the opportunity to research this

area during a period in which the practice of grading was still prevalent. Grading – along with the struggle so many of my peers identify - has offered an indicative externalization of what might, in the future, remain strictly internal: access to the Wisdom of Practice.

In my opinion, the most valuable and useful aspect of OTL is its collation of judgement-criteria, those strengths and improvements identified by observers in the classroom. These practical judgements - on their own and not used to ‘justify’ a grade – are the elements of ‘effective teaching’ that can be usefully converted to advice and guidance for the practitioner and feed-in to effective programmes of CPD, key points on an action-plan, or provide examples of good practice to share with others.

In the future, I believe that OTL should continue to play a major role in supporting ‘effective teaching’, and should be based on the ungraded judgements of experienced practitioners. The extent to which OTL will be valued by practitioners will be a consequence of the perceived expertise and skills of the observers who carry it out. Their judgements, based on the widely shared elements of ‘effective teaching’ that are contained in their Wisdom of Practice, constitute an invaluable resource – no matter how subjective and difficult to quantify they may be.

## References

- Abend, G. (2008), The meaning of 'theory'. *Sociological Theory*, 26 (2), pp. 173–199.
- Ainley, J. and Luntley, M. (2007). Towards an articulation of expert classroom practice, *Teaching and Teacher Education*, 23(7), pp. 1127-1138.
- Anderson, G. J. with Arsenault, N (1998). *Fundamentals of Educational Research*. Hove, UK: Psychology Press.
- Assessment Reform Group (2002). *Assessment For Learning: 10 principles*. Cambridge: Assessment Reform Group.
- Bailey, K. D. (1994). *Methods of Social Research*. New York: Free Press.
- Beishuizen, J., Hof, E., Van Putten, C., Bouwmeester, S. and Asscher, J. (2001). Students' and teachers' cognitions about good teachers, *British Journal of Educational Psychology*, 71(2), pp. 185-202.
- Berliner, D. C. (2005) The near impossibility of testing for teacher quality, *Journal of Teacher Education*, Vol. 56, (3), pp. 205-213.
- Bernard, H. R. (2011). *Research methods in anthropology*. California: Rowman Altamira.
- Black, P and Wiliam, D (1998). *Inside the Black Box: Raising standards through classroom assessment*. London: King's College.
- Black, P., Harrison, C. and Lee, C. (2003). *Assessment for learning: Putting it into practice*. London: McGraw-Hill Education.
- Boocock, A. (2013). Observation of teaching and learning: teacher development or micropolitics and neo-Fordism, *Journal of Further and Higher Education*, 37(4), pp. 482-503.

- Bogdan, R., and Biklen, S. K. (2007). *Qualitative Research For Education: An introduction to theory and methods*. Boston, MA: Pearson Allyn and Bacon.
- Boslaugh, S. (2012). *Statistics in a Nutshell: A desktop quick reference*. Sebastopol, Ca: O'Reilly Media.
- Bowling, A. (2005). Just one question: if one question works, why ask several?, *Journal of Epidemiology and Community Health*, 59 (5), pp. 342-345.
- Brown, S.; McIntyre, D. and McAlpine, A. (1988). The knowledge which underpins the craft of teaching, paper presented to the *Annual Meeting of the American Educational Research Association*, April 5-9th 1988, in New Orleans. [Online] Available at: <http://files.eric.ed.gov/fulltext/ED294872.pdf> [Accessed 4th October 2016].
- Brown, S. A. and McIntyre, D. (1993). *Making Sense Of Teaching*. Buckingham: Open University Press.
- Buchmann, M. (1987). Teaching knowledge: the lights that teachers live by, *Oxford Review of Education*, 13(2), pp. 151-164.
- Burnell, I. (2016). Teaching and learning in further education: the Ofsted factor, *Journal of Further and Higher Education*, 41 (2), pp. 227-237.
- Butler, R. and Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance, *Journal of educational psychology*, 78(3), p 210.
- Camus, A. (1975). *The Myth of Sisyphus*. London: Penguin.
- Carrette, J. (2007). *Religion and Critical Psychology: Religious experience in the knowledge economy*. London: Routledge.

Cercone, K. (2008). Characteristics of adult learners with implications for online learning design, *AACE journal*, 16(2), pp. 137-159.

Clarke, K. A. and Primo, D. M. (2007). Modernizing political science: a model-based approach, *Perspectives on Politics*, 5(4), pp. 741-753.

Cockburn, J. (2005). Perspectives and politics of classroom observation, *Research in Post-Compulsory Education*, 10(3), pp. 373-388.

Coe, R. (2014). *Classroom Observation: It's harder than you think*, [Online] Available at: <http://www.cem.org/blog/414/> [Accessed 12th January 2016].

Coe, R., Aloisi, C., Higgins, S. and Major, L. E. (2014). *What makes great teaching? Review of the underpinning research*. Sutton Trust, London.

Cohen, L., Manion, L. and Morrison, K. (2007). *Research methods in education*, 6th ed. London; New York: Routledge.

Coleman, M. (2012) Interviews, in Briggs, A; Coleman, M; & Morrison, M. (Eds) *Research Methods in Educational Leadership & Management*. Los Angeles: Sage

Creemers, B. P. and Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: the importance of establishing a dynamic model, *School Effectiveness and School Improvement*, 17(3), pp. 347-366.

Danielson, C. (2011). *Enhancing Professional Practice: A framework for teaching*. Alexandria, VA: ASCD.

Davidovitch, N. and Soen, D. (2006). Class attendance and students' evaluation of their college instructors, *College Student Journal*, 40(3), pp. 691–703.

- Denzin, N.K. (1969). Symbolic interactionism and ethnomethodology: a proposed synthesis. *American Sociological Review*, 34 (6), pp. 922-934.
- Department for Education and Skills (2001). *Adult Literacy Core Curriculum*. Nottingham: DfES Publications
- Department for Education and Skills (2004). *Equipping our Teachers for the Future: Reforming Initial Teacher Training for the Learning and Skills Sector*. London: DfES.
- Dewey, J. (1939) Theory of valuation, *Philosophy of Science*, 6(4), pp. 490-491.
- Dreyfus, S. E. and Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. University of California, Berkeley, [Online] available at: <http://www.dtic.mil/dtic/tr/fulltext/u2/a084551.pdf> [Accessed 14th November 2016].
- Dunne, R. and Wragg, E.C. (1994). *Effective Teaching*. London: Routledge
- Education and Training Foundation. (2014). *Initial Guidance for Users of the Education and Training Foundation for Teachers and Trainers in Education and Training – England*. London: EFT.
- Ellis, C., Adams, T. E. and Bochner, A. P. (2011). Autoethnography: an overview, *Historical Social Research/Historische Sozialforschung*, 12 (1), pp. 273-290.
- Fenstermacher, G. and Richardson, V. (2005). On making determinations of quality in teaching, *The Teachers College Record*, 107(1), pp 186-213.
- Finlay, L. (2008). Reflecting on ‘reflective practice’, in *PBLB paper 52*. Buckinghamshire: Open University Press

Galletta, A. (2013). *Mastering The Semi-Structured Interview and beyond: From research design to analysis and publication*. New York: NYU Press.

Garner, R. L. (2006). Humor in pedagogy: how ha-ha can lead to aha!, *College Teaching*, 54(1), pp. 177-180.

Gilroy, P. and Wilcox, B. (1997). Ofsted, criteria and the nature of social understanding: a Wittgensteinian critique of the practice of educational judgement, *British Journal of Educational Studies*, 45(1), pp. 22-38.

Glaser, B. G. and Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New Jersey: Transaction Publishers.

Gorden, R. L. (1998). *Basic Interviewing Skills*. Illinois: Waveland Press.

Gosling, D. (2002). Models of peer observation of teaching, *Generic Centre: Learning and Teaching Support Network*, [Online] Available at:  
[https://www.researchgate.net/profile/David\\_Gosling/publication/267687499\\_Models\\_of\\_Peer\\_Observation\\_of\\_Teaching/links/545b64810cf249070a7955d3.pdf](https://www.researchgate.net/profile/David_Gosling/publication/267687499_Models_of_Peer_Observation_of_Teaching/links/545b64810cf249070a7955d3.pdf)  
[Accessed 3rd April 2016].

Guba, E. G. and Lincoln, Y. S. (1994). Competing paradigms in qualitative research, *Handbook of Qualitative Research*, 2(163-194), pp 105-117.

Hargreaves, A. and Fullan, M. (2012). *Professional Capital: Transforming Teaching in Every School*. London: Routledge

Harper, H. (2013). *Outstanding Teaching in Lifelong Learning*. London: McGraw-Hill

Hastie, R. and Dawes, R. M. (2010). *Rational Choice in an Uncertain World : the psychology of judgment and decision making*, 2nd ed. Los Angeles: SAGE.



Hattie, J. (1999). Measuring the effects of schooling, *Australian Journal of Education*, 36(1), pp. 5-13.

Hawkins, D. (1966). Learning the unteachable, in: Shulman, L.S. and Keislar, E.R. (Eds) *Learning by Discovery: A critical appraisal*. Chicago: Rand McNally.

Heidegger, M. (1962). *Being and Time*. New York: Harper.

Hickman, R. D. (2009). *Cambridge Handbook of Educational Abbreviations and Terms*. Cambridge: University of Cambridge.

Higgs, J. (2012). Realising practical wisdom from the pursuit of wise practice, in Kinsella, E. and Pitman A. (Eds.), *Phronesis as Professional Knowledge: Practical Wisdom in the Professions*. Heidelberg: Springer Science and Business Media.

Hillage J., Uden T., Aldridge F., and Eccles J. (2000). *Adult Learning in England: a Review*. Brighton: Institute for Employment Studies.

Isaacs, T., Zara, C., Herbert, G., Coombs, S. J. and Smith, C. (2013). *Key Concepts in Educational Assessment*. Los Angeles: Sage.

James, D. and Biesta, G. (2007). *Improving Learning Cultures in Further Education*. Oxon: Routledge.

James, N. and Busher, H. (2016). Credibility, authenticity and voice: dilemmas in online interviewing, *Qualitative Research*, 6 (3), pp. 403 – 420.

Kaminski, M. (2004). *Games Prisoners Play: the tragicomic worlds of Polish prison*. Princeton, N.J.: Princeton University Press.

- Kinsella, E and Pitman, A. (2012). *Phronesis as Professional Knowledge: practical wisdom in the professions*. Heidelberg: Springer Science and Business Media.
- Ko, J. and Sammons, P. (2013). *Effective Teaching: A review of research and evidence*. Reading: CfBT Education Trust.
- Lasagabaster, D. and Sierra, J. M. (2011). Classroom observation: desirable conditions established by teachers, *European Journal of Teacher Education*, 34(4), pp. 449-463.
- LeCompte, M. D. and Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research, *Review of educational research*, 52(1), pp. 31-60.
- Leinhardt, G. (1990). Capturing craft knowledge in teaching, *Educational researcher*, 19(2), pp. 18-25.
- Lunenberg, M. and Korthagen, F. (2009). Experience, theory, and practical wisdom in teaching and teacher education, *Teachers and Teaching: theory and practice*, 15(2), pp. 225-240.
- Mathison, S., Ross, E. W., and Cornell, J. (1993). *Casebook for Teaching About Ethical Issues in Qualitative Research*. Washington: Qualitative Research SIG, American Educational Research Association.
- McBer, H. (2000). *A Model of Teacher Effectiveness*. London: DfEE.
- Mercer, J. (2007). The challenges of insider research in educational institutions: wielding a double-edged sword and resolving delicate dilemmas, *Oxford Review of Education*, 33(1), pp. 1-17.

- Mercer, J., Barker, B. and Bird, R. (2010). *Human Resource Management in Education: Contexts, themes and impact*. London: Routledge.
- Michaelidou, N. and Dibb, S. (2006). Using email questionnaires for research: good practice in tackling non-response, *Journal of Targeting, Measurement and Analysis for Marketing*, 14(4), pp. 289-296.
- Morgan, G and Smircich, L (1980). The case for qualitative research, *The Academy of Management Review*, Oct 1980; 5 (4), pp. 491-500.
- Muijs, D (2012). *Assessing Teacher Effectiveness: Different models*. London: Routledge.
- Muijs, D. and Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: some preliminary findings from the evaluation of the mathematics enhancement programme (primary), *School Effectiveness and School Improvement*, 11(3), pp. 273-303.
- Nixon, J. and Rudduck, J. (1993). The role of professional judgment in the local inspection of schools: a study of six local education authorities, *Research Papers in Education*, 8(2), pp. 135-148.
- Norris (1993). *Reconstructing Teacher Education: Teacher Development*. London: Falmer Press.
- Ofsted (1996). *Framework for the Assessment of Quality and Standards in ITT 1996/7*. London: Ofsted/TTA.
- Ofsted (1995). *The Guidance on Inspection of Secondary Schools*. London: Ofsted.
- Ofsted (2014). *Common Inspection Framework for Further Education and Skills*. London: Ofsted.

- Ofsted (2015). *Early Years Inspection Handbook*. London: Ofsted.
- O'Leary, M. (2006). Can inspectors really improve the quality of teaching in the PCE sector? Classroom observations under the microscope, *Research in Post-compulsory Education*, 11(2), pp. 191-198.
- O'Leary, M. (2011). *The Role of Lesson Observation in Shaping Professional Identity, Learning and Development in Further Education Colleges in the West Midlands*, PhD Thesis. Coventry: University of Warwick, Institute of Education.
- O'Leary, M. (2012). Exploring the role of lesson observation in the English education system: a review of methods, models and meanings, *Professional development in education*, 38(5), pp. 791-810.
- O'Leary, M. (2013). Surveillance, performativity and normalised practice: the use and impact of graded lesson observations in Further Education colleges, *Journal of further and higher education*, 37(5), pp. 694-714.
- O'Leary, M. (2013b). *Developing a National Framework for the Effective Use of Lesson Observation in Further Education*. London: UCU
- Pianta, R. C., K. M. La Paro, and B. K. Hamre. (2008). *Classroom Assessment Scoring System; Manual Pre-K*. Baltimore, MD: Brookes.
- Polanyi, M. (1962). Tacit knowing: its bearing on some problems of philosophy, *Reviews of modern physics*, 34(4), pp. 601-616.
- Provis, C. (2010). Virtuous decision making for business ethics, *Journal of Business Ethics*, 91(1), pp. 3-16.

Qu, S. Q. and Dumay, J. (2011). The qualitative research interview, *Qualitative Research in Accounting and Management*, 8(3), pp. 238-264.

Rosenshine, B. (2012). Principles of instruction: research-based strategies that all teachers should know, *American Educator*, 36(1), pp. 12-19.

Sadler, D. R. (1985) The origins and functions of evaluative criteria, *Educational Theory*, 35(3), 285-297.

Sartre, J-P (1948). *Existentialism and Humanism*. London: Methuen.

Schaefer, D. R. and Dillman, D. A. (1998). Development of a standard e-mail methodology: results of an experiment, *Public Opinion Quarterly*, 62(3), pp. 378-397.

Schwab, J. J. (1971). The practical: arts of eclectic, *The School Review*, 79 (4), pp. 493-542.

Seidman, I. (2006) *Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences*. New York: Teachers College Press.

Shulman, L. S. (1983). "Autonomy and obligation: The remote control of teaching." In: L. S. Shulman and G. Sykes (Eds.), *Handbook of Teaching and Policy*. New York: Longman.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform, *Harvard Educational Review*, 57(1), pp. 1-23.

Shulman, L. S. (1988). The paradox of teacher assessment, in Joanne Pfleiderer (Ed.) *New Directions for Teacher Assessment: Invitational conference proceedings*. Princeton, NJ: Educational Testing Service.

Shulman, L. S., Wilson, S. M. and Hutchings, P. (2004). *The Wisdom of Practice: Essays on teaching, learning, and learning to teach*. San Francisco: Jossey-Bass.

Sikes, P. (2004) Methodology, procedures and ethical concerns, in Opie, C. (Ed.) *Doing Educational Research*. Los Angeles: SAGE Publications Ltd.

Sikes, P. (2008). Researching research cultures, in: Sikes, P. and Potts, A. (Eds.), *Researching Education From The Inside: Investigations from within*. London: Routledge.

Sikes, P. and Potts, A. (2008). *Researching Education From The Inside: Investigations from within*. London: Routledge.

Smith, L. (1990). Ethics, field studies, and the paradigm crisis, in E. Guba (Ed.), *The Paradigm Dialog*. Newbury Park: Sage.

Smyth, A. and Holian, R. (2008). Credibility issues in research from within organisations, in Sikes, P. and Potts, A. (Eds.), *Researching Education From The Inside: Investigations from within*. London: Routledge.

Stobart, G. (2008) *Testing Times: The Uses And Abuses Of Assessment*. London: Routledge.

Strong, M., Gargani, J. and Hacifazlıoğlu, Ö. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers, *Journal of Teacher Education*, 62(4) pp. 367–382.

Stronge, J. H. (2007). *Qualities of Effective Teachers*. Alexandria, VA: ASCD.

Todres, L. and Wheeler, S. (2001). The complementarity of phenomenology, hermeneutics and existentialism as a philosophical perspective for nursing research, *International Journal of Nursing Studies*, 38(1), pp. 1-8.

Taylor, J. (2011). The intimate insider: negotiating the ethics of friendship when doing insider research, *Qualitative Research*, 11(1), pp. 3-22.

- Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D. and Yu, F. (2006). The international system for teacher observation and feedback: evolution of an international study of teacher effectiveness constructs 1, *Educational Research and Evaluation*, 12(6), pp. 561-582.
- Training and Development Agency for Schools (2007). *Annual Report and Accounts 2006–07*. Norwich: The Stationery Office.
- Turner, G. and Clift, P. (1988). *Studies in Teacher Appraisal*. Lewes: The Falmer Press.
- Van Selm, M. and Jankowski, N. W. (2006). Conducting online surveys, *Quality and Quantity*, 40(3), pp. 435-456.
- Wallace, M. and Hoyle, E. (2007) Educational reform an ironic perspective, *Educational Management Administration & Leadership*, 35(1) pp. 9–25
- Walsham, G. (2006) Doing interpretive research, *European Journal of Information Systems*, 15(3), pp. 320-330.
- Weimer, M. (2001). Learning more from the wisdom of practice, *New Directions for Teaching and Learning*, 86, pp. 45-56.
- White, F. C. (1992). Enhancing class attendance, *NACTA Journal*, 36 (4), pp.13-15.
- Wilcox, B. and Gray, J. (1996). *Inspecting Schools : Holding schools to account and helping schools to improve*. Buckingham: Open University Press.
- Winch, P. (1958). *The Idea of a Social Science and its Relation to Philosophy*. London: Routledge.
- Woods, P. and Jeffrey, B. (2002) The reconstruction of primary teachers' identities, *British Journal of Sociology of Education*, 23(1), pp. 89-106.

Wragg, E.C. (1999). *An Introduction to Classroom Observation*. London: Routledge.

Yeager, E. A. (2000). Thoughts on wise practice in the teaching of social studies, *Social Education*, 64 (6), pp. 352-353.



## **Appendices**

Appendix 1	Judgement-criteria cited in observation reports
Appendix 2	Auto-ethnography questionnaire items
Appendix 3	Interview schedule
Appendix 4	Elements of 'effective teaching' from literature sources
Appendix 5	Literature sources for elements of 'effective teaching'
Appendix 6	Ethical approval
Appendix 7	Observee questionnaire results for comparison purpose
Appendix 8:	Auto-ethnographic self-questionnaire (sample)

## Appendix 1: Judgement-criteria cited in observation reports

Judgement Criteria	Overall Mentions		As Strengths		As Improvements	
	frq	% occ	frq	% occ	frq	% occ
Differentiation	88	64	48	35	40	29
Feedback	70	51	38	28	32	23
Planning	70	51	47	34	23	17
ITC	65	47	42	31	23	17
Activity	63	46	52	38	11	8
Questioning technique	63	46	41	30	22	16
Classroom management	60	44	44	32	16	12
Pace	49	36	21	15	28	20
Resources	49	36	46	34	3	2
Teaching methods	49	36	16	12	33	24
Subject content	46	34	32	23	14	10
Aims	45	33	20	15	25	18
Peer working	39	28	25	18	14	10
Reflective practice	36	26	19	14	17	12
Paperwork	33	24	22	16	11	8
Accreditation focus	31	23	31	23		0
Clear instructions	31	23	8	6	23	17
E&D	31	23	25	18	6	4
Assessment	30	22	20	15	10	7
Engagement	29	21	27	20	2	1
Evidence of learning	26	19	20	15	6	4
Links to previous learning	22	16	22	16		0
Rapport	19	14	19	14		0
Managing volunteer	18	13	11	8	7	5
Plenary	18	13	12	9	6	4
Attendance	16	12	10	7	6	4
Missed opportunities	16	12			16	12
Checking of learning	15	11	9	7	6	4
Learning outside class	15	11	15	11		0
Knowledge of Ls' needs	14	10	14	10		0
Enthusiastic delivery	13	9	13	10		0
Independent learning	13	9	13	10		0
Learner voice	12	9	10	7	2	1
Challenging work	8	6	4	3	4	3
Learning environment	8	6	7	5	1	1
Accreditation rate	5	4	4	3	1	1
Relevance	3	2	3	2		0
Active Learning	2	1			2	1
Learner centred	2	1	2	1		0
Competition element	1	1	1	1		0
Employability focus	1	1	1	1		0
Excellent external visit	1	1	1	1		0
Familiar with Ls' names	1	1	1	1		0
Give learners complete T/L	1	1			1	1
Guidance & support	1	1	1	1		0
Learners' work	1	1	1	1		0
Ls' aware of learning journey	1	1	1	1		0
Ls clear about progress	1	1	1	1		0
Obscure meta-language	1	1			1	1
Progression into work	1	1	1	1		0
Punctuality of Ls	1	1			1	1
Sensitivity to using L's work	1	1			1	1
Speaking to Ls appropriately	1	1			1	1
Tutor's handwriting	1	1			1	1
Valuing Ls' contributions	1	1			1	1
Venue issue	1	1			1	1

Table A1.1: Full list of judgement-criteria cited in observation reports

## **Appendix 2: Auto-ethnography questionnaire items**

### **A: Prior to observation**

1. What are my initial feelings about the experience to come?
2. How will I prepare for the observation?
3. Will I do anything differently than I would have done if I were not being observed?
4. How do I feel as the date of observation approaches?
5. What are my priorities for the observed session?
6. Do these priorities differ from those of a standard teaching session? If so, how?
7. How well prepared do I feel for the observed session?
8. Do I feel differently about the impending session?
9. How do I feel about the observer/observers before the session

### **B: After the observation**

10. Reflect on the experience of being observed
11. How do you feel now?
12. Assess whether extra time spent preparing was “worth it”
13. How do you feel you performed?
14. Did you behave as you normally do in a teaching session?
15. Where was your focus during the observed session?
16. How do you think the learners were affected by the observation?

### C: Result and Feedback

17. How was the result of the observation delivered?
18. How do you feel about the result?
19. Do you agree or disagree with the feedback given?
20. Is the feedback useful to you?
21. How will the feedback affect your future teaching?
22. How do you now feel about your observers?

### Appendix 3: Interview schedule

Group	Question	Prompt
<b>Purpose &amp; Value</b>	What do you classroom observation is for?	Any other purposes?
	Do you think it is important?	To whom?
		Why?
	Is it as good as it could be at the moment, or could it be better?	In what ways?
<b>Tensions</b>	How do you feel generally about the process of observing or being observed?	Being Observed
		Observing
		Do any elements arising affect your decision process?
	Does the length of experience of the tutor make it easier or more difficult for the observer?	How?
		Why?
		Examples
	What would you do in a case when the observation was really border-line between grades?	Some people I've spoken to say it is more difficult to observe someone you know well...
<b>Decision Making</b>	What do you base your judgements on?	Checklists?
		Training?
		Ofsted?
	What are the most important things you look for in an observed lesson?	Types of skills
		Behaviours
		Environment
		External factors?
	Do you use any formal or informal checklist of strengths and weaknesses?	Are you aware of being influenced by any training or teaching frameworks?
<b>Finally...</b>	How do you personally decide on an observation grade?	Where does it come from?
		How does it coalesce?
		What is your thought process?

## Appendix 4: Elements of ‘effective teaching’ from literature sources

Note: Elements listed using the same nomenclature as for judgement-criteria.

Element of Effective Teaching	Occurance in sources
Assessment	21
Planning	20
Feedback	19
Learning environment	19
Activity	18
Aims	18
Challenging work	18
Clear instructions	18
Engagement	18
Subject content	18
Differentiation	17
Checking of learning	16
Classroom management	16
Questioning technique	16
Reflective practice	16
Teaching methods	16
Independent learning	15
Guidance & support	14
Knowledge of Ls' needs	14
Resources	13
Ls' aware of learning journey	12
Rapport	12
Enthusiastic delivery	11
Pace	11
Evidence of learning	10
Learner centred	10
Learning outside class	10
Links to previous learning	10
Peer working	10
Ls clear about progress	9
Plenary	9
Relevance	9
E&D	8
Learners' work	8
ITC	7
Venue issue	7
Discipline	7
Speaking to Ls appropriately	6
Valuing Ls' contributions	6
metacognitive skills	6
Active Learning	5
Learner voice	5
Managing volunteer	5
Accreditation rate	4
Attendance	4
Punctuality of Ls	4
Missed opportunities	3
Obscure meta-language	3
Paperwork	3
Progression into work	3
Employability focus	2
Familiar with Ls' names	2
Relationships with colleagues	2
teacher beliefs	2
Accreditation focus	1
Excellent external visit	1
Sensitivity to using L's work	1
Tutor's handwriting	1
flexibility	1
L's Self-confidence	1
evidence-based practice	1

Table A4.1: Full list of elements of ‘effective teaching’ from literature sources

## Appendix 5: Literature sources for elements of ‘effective teaching’

<b>Governmental/Agency guidelines and advisory documentation</b>
Department for Education and Skills (2004)
Training and Development Agency for Schools (2007)
The Education and Training Foundation (2014)
Ofsted (2014)
<b>Professional literature/guidance</b>
Dunne & Wragg (1994)
ARG (2002)
Teddlie et al. (2006)
CLASS (Pianta et al., 2008)
Danielson, C (2011)
Coe et al (2014)

Academic studies
Shulman (1987)
Brown & McIntyre (1993)
Black & Wiliam (1998)
Hattie (1999)
McBer (2000)
Muijs and Reynolds 2000
Beishuizen et al (2001)
Black et al. (2003)
Stronge (2007)
Kyriacou (2007)
Rosenshine (2012)
Creemers & Kyriakides (2006)
Harper (2013)
Ko & Sammons (2013)

*Table A5.1: Sources accessed in the survey of literature on the subject of 'effective teaching'*



## Appendix 6: Ethical approval

Notes: This form is abridged in order to save space, and because the original held details of the research setting that might compromise participant confidentiality. All details of the text of the application are contained within the relevant chapters above.

Condition 1 (below): Confirmation that my laptop was password protected was given and accepted

Centre for Education Studies

### Application for Ethical Approval for Research Degrees

(MA by research, MPHIL/PhD, EdD)

Name of student

EdD

**Graeme Robert Sutherland**

Project title: Can the current practice of the Observation of Teaching and Learning (OTL) be considered 'useful' within the context of Adult Community Education?

Supervisor: **Dr Michael Hammond / Dr Justine Mercer**

Please ensure you have read the Guidance for the Ethical Conduct of Research available in the handbook.

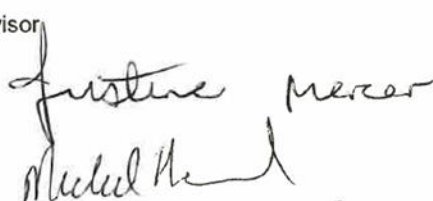
In reference to the attribution of authorship by yourself and your supervisor(s) of any reports or publications, I will follow the procedures as stated in the research student handbook.

Research student



Date 29/10/14

Supervisor



Date

29/10/14  
29/10/14

Action taken

☐

Approved

☒

Approved with modification or conditions – see below

☐

Action deferred. Please supply additional information or clarification – see below

① Please confirm that your laptop is password protected ✓

Name

G. Hindley

Date

30/10/14

Signature



Stamped

THE UNIVERSITY OF  
WARWICK

Notes of Action

CENTRE FOR EDUCATION STUDIES

## Appendix 7: Observee questionnaire results for comparison purpose

[The following table is included for interest. The work it pertains to was excluded from the final thesis as explained in section 7.5, p196]

21 Essential Skills tutors (observees) were asked to complete a single-item questionnaire on 10/09/15. They were asked “What do you consider the most important skills or elements of good classroom teaching?” The table below compares the most frequent elements mentioned by the observees with most frequent judgement-criteria mentioned by their observers in the observation reports.

<b>Mentioned by Observees</b>	<b>frq</b>	<b>Mentioned in Observation Reports (rank order)</b>
Engagement	14	Differentiation
Aims	14	Feedback
Differentiation	12	Planning
Learning environment	10	ICT
Resources	9	Activity
Rapport	8	Questioning technique
Personal Qualities of Tutor	8	Classroom management
Independent learning	7	Pace
Evidence of learning	6	Resources
Teaching methods	6	Teaching methods
Activities	6	Subject content
Assessment	6	Aims
Subject content	6	Peer working
ICT	6	Reflective practice
Relevance	5	Paperwork
Classroom Management	5	Accreditation focus
Checking of learning	5	Clear instructions
Peer working	5	E&D
Questioning technique	5	Assessment
Knowledge of L's needs	5	Engagement

*Table A7.1: Comparison between the 20 most frequent elements of 'effective teaching' from the observee questionnaire with most frequent judgement-criteria in the observation reports*